Cross-section measurement of top quark pair production with additional b-jets in the $e\mu$ channel at $\sqrt{s}$ = 13 TeV with the ATLAS detector

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium (Dr. rer. nat.)

im Fach Physik

Spezialisierung: Experimentalphysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von

M.Sc. Akanksha Vishwakarma

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr.-Ing. Dr. Sabine Kunst

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät
Prof. Dr. Elmar Kulke

Gutachter/innen:
1. Prof. Dr. Klaus Mönig
2. Prof. Dr. Ramona Gröber
3. Prof. Dr. Thomas Lohse

Tag der mündlichen Prüfung: 23.09.2019

Zusammenfassung

Diese Doktorarbeit behandelt die Messung des Wirkungsquerschnittes der Paarbildung von Top Quarks mit zusätzlichen B-Jets (Teilchenschauer aus der Fragmentierung von B-Quarks) im Proton-Proton-Kollisionen bei einer Schwerpunktsenergie von $\sqrt{s} = 13\,\text{TeV}$ am Large Hadron Collider. Der Datensatz wurde 2015/16 mit dem ATLAS Detektor aufgezeichnet und entspricht einer integrierten Luminosität von $36.1\,\text{fb}^{-1}$. Es wurden Top-Paar-Ereignisse mit je einem Elektron, einem Myon und mindestens zwei B-Jets im Endzustand untersucht. Zur Minimierung von modellabhängigen Korrekturen wird der Wirkungsquerschnitte in einem fiduziellen Phasenraum bestimmt, der nah an der Detektorakzeptanz der Physikobjekte gewählt wurde. Die Ergebnisse werden auf Teilchenebene als inklusiven und differenzielle Wirkungsquerschnitte von $t\bar{t}$ Endzutänden mit mindestens einem ($t\bar{t}b$) oder zwei ($t\bar{t}b\bar{b}$) zusätzlichen B-Jets präsentiert. Die Messungen werden mit dem neusten Stand von unterschiedlichen theoretischen Vorhersagen verglichen.

Die gemessenen inklusiven, fiduziellen Wirkungsquerschnitte in $t\bar{t}b$- und $t\bar{t}b\bar{b}$-Endzuständen sind:

$$\begin{aligned} \sigma_{t\bar{t}b} &= 181 \pm 5(\text{stat}) \pm 24(\text{syst})\ \text{fb} \\ \sigma_{t\bar{t}b\bar{b}} &= 27 \pm 3(\text{stat}) \pm 7(\text{syst})\ \text{fb} \end{aligned}$$

Die QCD-Produktion von zusätzlichen B-Jets, die durch Abzug der Beiträge von $t\bar{t}X(b\bar{b})$- wobei X entweder für ein Z- oder Higgs-Boson steht - erhalten werden, kann durch

$$\begin{aligned} \sigma_{t\bar{t}b}^{\text{Data}-t\bar{t}X} &= 177 \pm 5(\text{stat}) \pm 24(\text{syst})\ \text{fb} \\ \sigma_{t\bar{t}b\bar{b}}^{\text{Data}-t\bar{t}X} &= 25 \pm 3(\text{stat}) \pm 7(\text{syst})\ \text{fb} \end{aligned}$$

abgeschätzt werden.

Im Allgemeinen sind die gemessenen inklusiven, fiduziellen Wirkungsquerschnitte größer als die $t\bar{t}b\bar{b}$-Vorhersagen von verschiedenen NLO-Berechnungen des mit Partonschauern abgeglichenen $t\bar{t}$-Matrixelements, bleiben aber im Rahmen der Messunsicherheit. Die Unsicherheiten der experimentellen Messung sind kleiner als die Unsicherheiten der Vorhersagen.

Die differenzielle Messung von $t\bar{t}$ + B-Jets wird mit dieser Analyse zum ersten Mal innerhalb der ATLAS-Kollaboration durchgeführt. Die normalisierten differenziellen Wirkungsquerschnitte werden als Funktion der Multiplizität der B-Jets, kinematischen Ereignisobservablen und Observablen der Objekte in den Endzuständen dargestellt. Die beobachteten differenziellen Verteilungen werden mit einer Vielzahl an theoretischen Vorhersagen verglichen und eine gute Übereinstimmung mit den Daten für die meisten festgestellt.

Abstract

This thesis presents the cross-section measurement of top quark pair ($t\bar{t}$) production in association with additional b -jets ($t\bar{t}$ + b-jets) in proton-proton collisions at a centre-of-mass energy $\sqrt{s}$ = 13 TeV at the Large Hadron Collider. The analysis is performed in the $e\mu$ decay channel of $t\bar{t}$ using a dataset comprising an integrated luminosity of 36.1 fb^{-1} collected by the ATLAS detector in the years 2015 and 2016. The selected events consist of one electron, one muon and at least two b-jets. To minimize the corrections from model dependence the cross-sections are measured in a fiducial phase space which is chosen close to the detector acceptance of the physics objects. Results are presented at particle level in the form of inclusive and differential cross-sections of $t\bar{t}$ final states with at least one ($t\bar{t}b$) or two ($t\bar{t}b\bar{b}$) additional b-jets. Comparisons of the measurements to various state-of-the-art theoretical predictions are made.

The measured fiducial inclusive cross-section in $t\bar{t}b$ and $t\bar{t}b\bar{b}$ final states are

$$\begin{aligned} \sigma_{t\bar{t}b} &= 181 \pm 5(\text{stat}) \pm 24(\text{syst}) \text{ fb} \\ \sigma_{t\bar{t}b\bar{b}} &= 27 \pm 3(\text{stat}) \pm 7(\text{syst}) \text{ fb} \end{aligned}$$

Estimates of the QCD production of additional b-jets obtained after subtracting the contribution from $t\bar{t}X(b\bar{b})$ where X can be either Z or Higgs boson are

$$\begin{aligned} \sigma_{t\bar{t}b}^{\text{Data}-t\bar{t}X} &= 177 \pm 5(\text{stat}) \pm 24(\text{syst}) \text{ fb} \\ \sigma_{t\bar{t}b\bar{b}}^{\text{Data}-t\bar{t}X} &= 25 \pm 3(\text{stat}) \pm 7(\text{syst}) \text{ fb} \end{aligned}$$

The measured inclusive fiducial cross-sections generally exceed the $t\bar{t}b\bar{b}$ predictions from various next-to-leading-order matrix element calculations matched to a parton shower but are compatible within the uncertainties. The experimental uncertainties in the measurement are smaller than the uncertainties in the predictions.

The differential measurement of $t\bar{t}$ + b-jets is carried out for the first time by the ATLAS collaboration in this analysis. The normalised differential cross-sections as a function of b-jet multiplicity, kinematic observables of the events and properties of b-jet pairs are shown. The observed differential distributions are compared with a multitude of theoretical predictions and a good agreement with data is found for most of them.

Declaration

I declare that I have completed the thesis independently using only the aids and tools specified. I have not applied for a doctor's degree in the doctoral subject elsewhere and do not hold a corresponding doctor's degree. I have taken due note of the Faculty of Mathematics and Natural Sciences PhD Regulations, published in the Official Gazette of Humboldt-Universität zu Berlin no. 42/2018 on 11/07/2018.

Date: Signature

Place:

Akanksha Vishwakarma

Acknowledgements

I express my profound gratitude to my supervisor Prof. Dr. Klaus Mönig for giving me the opportunity to join his research group and opening the possibilities of various research areas where I contributed. His invaluable guidance and constant supervision nurtured my doctorate journey. I am lacking words to express my thankfulness for his patience.

I am extremely indebted to Dr. Thorsten Kuhl for indulging into discussions throughout my research work. His critical remarks and suggestions have always been very helpful in gaining comprehensive knowledge, especially during the thesis writing. I would like to express my sincere gratitude to Dr. Judith Katzy for always being the source of motivation for me and having faith in me during the hardships. Every meeting with her boosted my confidence.

I convey my deep acknowledgement to Dr. Mahsana Haleem for her priceless mentorship, friendly approach in guiding me during this analysis and working for late hours together. I would take the opportunity to express my special appreciation to Dr. Baishali Dutta for always clarifying my tiniest doubts and reviewing this thesis. Many thanks to Yasiel for fruitful discussions and help in the analysis. Thanks to Matthieu, Filip, Wrijupan and all my colleagues at DESY for providing a vibrant academic environment as well as fun moments.

Most importantly, it gives me a great pleasure in expressing my heartfelt gratitude to my parents for their endless support and motivation throughout my education. None of this would have been possible without them. I am also grateful to my brother and sister who were always there for me despite being away. I feel blessed to have so supportive in-laws who always find new ideas and stories to inspire me. A special mention to my dearest friend Shalini for listening to my plight over endless calls and uplifting my courage. Above all, my immense thanks to my fiancé Naveen for his encouragement, admiration and moral support which has provided a smooth basis during the whole phase of PhD. His presence made this challenging journey joyous.

Contents

Introduction

The search for a description of the most fundamental building blocks of the universe is perhaps as old as the history of humankind. The successful model which explains the elementary particles of the visible universe and their interactions, except gravitation, is known as the Standard Model (SM) of particle physics. Developed in mid of 1970s, it explained almost all the experimental results obtained by then and precisely predicted a wide variety of phenomena. It received major acceptance by the discovery of myriad of particles predicted by it, with the advent of experiments. The Large Hadron Collider (LHC) at CERN with its unprecedented collision energy successfully discovered the Higgs boson [1, 2], establishing the last missing piece of the SM predictions. The immensely high proton-proton collision energy at LHC allows the precision measurements and test of the SM and the discovery of new phenomena.

One of the most intriguing areas of particle physics is the top quark sector. Being the heaviest among all the known elementary particles, the top quark possesses unique properties such as its strong coupling with the Higgs boson. Its large mass also indicates coupling with *new* particles predicted in physics beyond the standard model (BSM) providing sensitivity to new phenomena. Several BSM models [3–7] involve the decay of new heavy resonances into top quarks or production of new particles in association with it. The production of top quark in abundance at LHC gives the opportunity to study its properties in greater details.

The measurement of top-antitop quark pair ($t\bar{t}$) production in association with jets provides stringent tests of the Quantum Chromodynamics (QCD) predictions. Among these, the process involving $t\bar{t}$ production with additional jets originating from b-quarks ($t\bar{t} + b$-jets) is of particular importance. The presence of two QCD mass scales (top and bottom quarks) and multi-particle complex final state induce several challenges for theoretical predictions, which suffer from large uncertainties. It is therefore important to provide experimental measurements of $t\bar{t} + b$-jets in order to compare the predictions with observed data to understand multi-scale QCD calculation and improve the theoretical modelling. Moreover, the direct investigation of Higgs coupling to the top quark ($t\bar{t}H$) [8, 9] is a vital test for better understanding of the origin of mass. The $H \to b\bar{b}$ decay mode constitutes the highest branching ratio [10, 11] and therefore $t\bar{t}H$ can be measured with the best statistical precision when the Higgs boson decays to $b\bar{b}$. The $t\bar{t}H(H \to b\bar{b})$ process leads to the $t\bar{t}b\bar{b}$ final state and suffer from large irreducible

background coming from QCD induced $t\bar{t}b\bar{b}$ production.

Several measurements are published in the context of $t\bar{t}$ + b-jets by both ATLAS and CMS which are the two general purpose experiments at LHC. The ratio of cross-section of $t\bar{t}$ production with heavy flavour quarks to the cross-section of $t\bar{t}$ with one additional jet regardless of the flavour has been investigated by ATLAS at $\sqrt{s}$ = 7 TeV [12]. At 8 TeV, ATLAS presented the fiducial inclusive cross-section for $t\bar{t}$ production with one or two additional b-jets [13]. CMS has reported the fiducial inclusive cross-sections at 8 and 13 TeV [14–16] and in addition, the differential measurement as well at 8 TeV [16] which is dominated by the statistical uncertainties.

With the increased centre of mass energy and luminosity, this thesis is set out to measure the first ATLAS differential cross-section of $t\bar{t}$ production with at least one ($t\bar{t}b$) or two ($t\bar{t}b\bar{b}$) additional b-jets and also the inclusive fiducial cross-sections [17]. This analysis is performed using 36.1 fb^{-1} of proton-proton collision data collected by the ATLAS detector at $\sqrt{s}$ = 13 TeV. The selection of $e\mu$ final state for this analysis provides a clean signal and minimizes the possibilities of jet flavour mis-assignments. It is essential to measure the differential cross-section in addition to the inclusive cross-section to provide a detailed description of the process and test the predictions. The differential cross-section is measured as a function of b-jet multiplicity, p_{T} of the b-jets, global kinematic properties of the events such as the scalar sum of the p_{T} of all final state objects. The cross-sections are also measured as a function of the kinematic distributions of the b-jet pairs which provide substantial sensitivity to the contribution of specific production mechanisms.

Monte Carlo generators for the $t\bar{t}$ + b-jets final state developed vastly in the last years getting more predictive power. These MC models utilizes different approaches in the precision calculations which motivates their comparison with the observed data and assess the level of agreement between them. The mass of b-quarks which is very small as compared to the top quark but significant as compared to other light quarks allows the treatment of b-quarks as a massless or massive particle in the simulation. The models also allow the calculation of the additional b-jets in the matrix element or coming from the parton shower. All these criteria are dealt in different ways by different MC models. Along with the pre-existing MC models, the results obtained in this analysis are compared with the latest available theoretical predictions. State-of-the-art QCD calculations give predictions for the $t\bar{t}$ production cross-section with up to one additional massless partons at next-to-leading order (NLO) in perturbation theory matched to a parton shower [18] and the QCD production of $t\bar{t}b\bar{b}$ in the matrix element is calculated at NLO matched to a parton shower [19–21]. A simulation based subtraction of $t\bar{t}X$ ($X = H, Z$) is also performed in order to compare the results directly with the QCD calculations. This analysis

provides the data required for the improvement of the MC simulations of the $t\bar{t} + b$-jets process.

Chapter 1 gives the theoretical overview of the $t\bar{t} + b$-jets process in the context of the SM and different aspects of proton-proton collision phenomenology. It also discusses the production and decay of $t\bar{t}$ and different sources of production of additional b-jets. In Chapter 2, the description of the LHC accelerator, the ATLAS detector and its subsystems is provided with particular emphasis on the dataset used in this analysis. Chapter 3 elaborates the different steps and approaches used to simulate the $t\bar{t} + b$-jets process and lists all the MC simulated samples used for comparison. Chapter 4 explains the algorithms used to obtain the reconstructed objects and provides the definitions of generated particle-level objects in the MC simulations.

Chapter 5 provides the details of signal selection for the $t\bar{t} + b$-jets process and estimation of all the contributing backgrounds. Along with the non-$t\bar{t}$ background, this analysis also handles $t\bar{t}$ related background coming from the charm and light-jets which are misidentified as b-jets using a data-driven template fit. A data-MC comparison is shown at the detector-level.

The cross-section measurement is presented in Chapter 6. It is important to obtain the results independent of the detector specifications in order to facilitate the comparison with MC models. The measurements are performed at particle-level by correcting the data for the detector effects using an unfolding procedure in a fiducial phase space that is defined very close to the detector acceptance volume.

Chapter 7 presents the results of fiducial inclusive and normalised differential cross-sections compared with various MC predictions. The conclusion and outlook of this work is outlined in Chapter 8.

Chapter 1.

Theoretical Basis

This chapter provides an overview of the theoretical basis required to understand the top quark pair ($t\bar{t}$) and associated bottom quark production processes in the context of Standard Model (SM) summarised in Section 1.1. The SM of particle physics is a well established theory so far that describes the elementary particles and the interactions between them. The theory has evolved gradually over many decades in the 20^{th} century and finalized the current formulation in 1970s. Several searches and precision measurements performed with the development of large experiments exhibited a very good agreement with the predictions of the SM.

One such stringent test of the SM is the cross-section measurement of $t\bar{t}$ production with additional bottom quarks, as performed in this analysis using the dataset provided by the proton-proton (pp) collisions at the Large Hadron Collider (LHC). The phenomenology of the pp collision is discussed in Section 1.2. In Section 1.3, the production and decay processes of the top quark pair are described. Section 1.4 discusses the origin of additional b-quarks and the challenges arising in the study of such a complex process.

1.1. The Standard Model of Particle Physics

The mathematical formulation of the SM is based on quantum field theory (QFT) which describes collectively the electromagnetic, weak and strong interactions between the fundamental constituents of matter [22–26]. The physical matter is constituted by elementary particles called *fermions* and the interactions between them are explained in terms of the exchange of *gauge bosons*. Section 1.1.1 details the characteristics of these elementary particles.

The fundamental notion of symmetries is the basis of the Standard Model. As inferred from the Noether's theorem [27], with each continuous symmetry of the Lagrangian there is an associated conservation law. The global symmetries like the invariance of the Lagrangian

under space, time translation and rotation give rise to the conservation of the total momentum in absence of any external force, conservation of energy and conservation of the total angular momentum, respectively. The SM also possesses internal symmetries which act at subatomic scales leading to the associated conservation laws in particle physics.

In QFT approach, the fermions are recognised as local excitations of the underlying fermion fields. The interaction between these fundamental fields are introduced as the gauge fields and the excitations of these fields are called the gauge bosons. The Lagrangian density of the interaction field is required to be invariant under the local space-time transformation of the fields. This is known as the *gauge invariance*. The combined internal symmetry group of the SM is represented as

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{1.1}$$

where the strong interaction is described by the $SU(3)_C$ group, i.e. a group of 3×3 special unitary matrices and its conserved quantum number is called *colour*. The subscript L in the $SU(2)_L$ implies that the symmetry applies to left-handed fermion[1] fields and the conserved quantity is *weak isospin*, T. The weak interaction is based on $U(1)_Y$ symmetry group, which is a group of complex numbers with unitary module. It has one conserved quantum number defined as *weak hypercharge*, $Y \equiv 2(Q - T)$ where Q stands for electric charge. The strong interaction is formulated in quantum chromodynamics (QCD) detailed in Section 1.1.3 whereas the the electromagnetic and weak interactions are described in electroweak (EW) theory (Section 1.1.2). The SM Lagrangian $\mathcal{L}_{SM}$ can be expressed as $\mathcal{L}_{SM} = \mathcal{L}_{QCD} + \mathcal{L}_{EW}$.

It should be noted that the SM provides remarkable insights into the fundamental constituents of matter and unifies the electromagnetic, weak and strong interactions in a single theory. However, it does not incorporate the gravitational force. The gravitational force is directly proportional to mass therefore its impact on the fundamental particles having extremely low mass is very small. Hence, it becomes a negligible contributor to the interactions of fundamental particles as compared to the other three forces, which allows gravity to be ignored in the SM.

1.1.1. Elementary Particles

Fermions are half-integer spin particles which obey Fermi-Dirac statistics [28, 29] and Pauli exclusion principle [30]. These are further categorised into two groups namely *quarks* and *leptons*. Each particle has its own antiparticle with same properties (like mass) but opposite

[1] The property called handedness or helicity is the same as chirality for the massless particles. If the spin of a particle is opposite to its direction of motion, it is left-handed.

quantum numbers (like electric charge). Both categories have three generations where the particles are arranged depending on increasing masses and the properties based on the mass, such as lifetime. The different types of quarks and leptons are also referred to as *flavours* e.g. there are six flavours of quarks.

- Each of the three generations of the leptons include two particles - a charged particle and its corresponding neutral particle. The charged particles electron, muon and tau carry electrical charge of -1q_e (q_e is the charge of proton) and the electron-neutrino, muon-neutrino and tau-neutrino are electrically neutral. Being neutral, the neutrinos interact only via weak interaction while other leptons being charged also interact via electromagnetic interaction.

- The generations of quarks are subdivided into 'up-type' and 'down-type' quarks with the electrical charge of +2/3q_e and -1/3q_e respectively. Up-type quarks consist of the up, charm and top quarks whereas the down-type comprises of down, strange and bottom quarks. In addition to the electric charge the quarks also carry colour charge arbitrarily chosen to be red, green or blue. Therefore, quarks interact with strong interaction along with the electromagnetic and weak interactions. Unlike leptons, the quarks cannot exist in isolation. Only colour neutral particles exist freely in nature which are called hadrons - mesons and baryons. The mesons are composed of a quark of any colour and an antiquark with the negative of that colour. The baryons, e.g. protons, are composition of three quarks with final colour-neutral state.

The properties of fermions are listed in Table 1.1. The neutrinos are considered to be massless in the SM but the observed neutrino oscillation [31] indicates their non-zero masses.

Fermions	Generation	Particle	Symbol	Mass	Electric charge
Leptons	1st	Electron	e	0.51 MeV	-1
		Electron neutrino	ν_e	<2 eV	0
	2nd	Muon	μ	105.66 MeV	-1
		Muon neutrino	ν_μ	<2 eV	0
	3rd	Tau	τ	1.77 GeV	-1
		Tau neutrino	ν_τ	<2 eV	0
Quarks	1st	Up	u	$2.2^{+0.5}_{-0.4}$ MeV	+2/3
		Down	d	$4.7^{+0.5}_{-0.3}$ MeV	-1/3
	2nd	Charm	c	$1.275^{+0.025}_{-0.035}$ GeV	+2/3
		Strange	s	95^{+9}_{-3} MeV	-1/3
	3rd	Top	t	173.0 ± 0.4 GeV	+2/3
		Bottom	b	$4.18^{+0.04}_{-0.03}$ GeV	-1/3

Table 1.1.: Summary of quarks and leptons with their properties in the Standard Model [32].

The gauge bosons are integer spin particles that obey Bose-Einstein [33,34] statistics.

- Photons are responsible for the electromagnetic interaction between the electrically charged particles. This interaction can either be repulsive or attractive depending on the charge of the particle and has infinite range.
- The $W^{\pm}$ and Z^0 bosons are responsible for the weak interaction which is exhibited by all fermions. These bosons have non-zero masses which are quoted in Table 1.2 with other bosons and their properties as in the SM. This interaction is short range and weaker as compared to the electromagnetic interaction. The unique property of the weak interaction is the flavour changing process, like the flavour transitions of the quarks. This is the only force interacting with neutrinos in the SM formalism.
- Gluons couple the quarks and are the mediator of strong interaction. These are massless and exist in a colour octet-state.

Gauge boson	Interaction	Mass [GeV]	Electric charge	Color
Photon (γ)	Electromagnetic	0	0	–
$W^{\pm}$	Weak	80.379 ± 0.012	± 1	–
Z^0	Weak	91.188 ± 0.0021	0	–
Gluon (g)	Strong	0	0	8 combinations
Higgs boson	–	125.09 ± 0.24	0	–

Table 1.2.: Summary of the gauge bosons with their associated interactions and properties. The Higgs boson of the SM is also included [32].

The discovery of the Higgs boson in 2012 [1,2] completed the list of particles in the SM. It is not a gauge boson that mediates a force, instead it is a boson which gives mass to the fundamental particles. Figure 1.1 is a nice illustration of the fermions, bosons and their interactions. The self-interaction of the bosons are indicated where applicable.

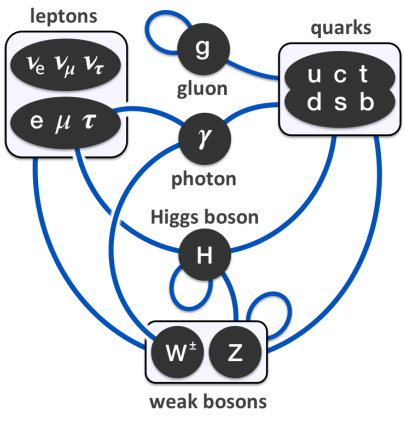


Figure 1.1.: Diagram summarizing the interactions between the elementary particles [35].

1.1.2. Electroweak Sector and Spontaneous Symmetry Breaking

The electromagnetic (EM) interaction is described by the most precisely known gauge field theory – Quantum Electrodynamics (QED) with $U(1)_{EM}$ Abelian symmetry group [36, 37]. In QED, the EM field is described as a field that couples to a charged particle via the exchange of photons. The strength of the interaction between the charged particles and photons is given by a small dimensionless EM coupling constant α.

$$\alpha \approx \frac{1}{137} \ll 1 \tag{1.2}$$

This value is valid at zero momentum transfer and the variation of α with the momentum transfer Q^2 is extremely small, for example at the scale of the Z boson mass ($Q = m_Z = 91.2\,\text{GeV}$) $\alpha(m_Z^2) \approx \frac{1}{128.957}$ [38].

The weak interaction is explained in the theory of Quantum Flavourdynamics [39] but it can be better understood in terms of electroweak theory.

The electromagnetic and weak interactions are entirely separate phenomena at low energies but at a scale of $\mathcal{O}(100\,\text{GeV})$ and higher both can be described by a single unified theory called the electroweak (EW) theory [40]. This theoretical model was given by Glashow, Weinberg and Salam [22, 41, 42] describing the $SU(2)_L \times U(1)_Y$ symmetry together as electroweak sector. The $SU(2)_L$ and $U(1)_Y$ corresponds to four generators which corresponds to four massless gauge bosons: $W_\mu^a (a = 1, 2, 3)$ and B_μ respectively where μ represents the Lorentz indices. The $\mathcal{L}_{EW}$ is given as

$$\mathcal{L}_{EW} = -\frac{1}{4} W_{\mu\nu}^a W_a^{\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} + \sum_{fermions} (\imath \bar{\psi}_L D_\mu \gamma^\mu \psi_L + \imath \bar{\psi}_R D_\mu \gamma^\mu \psi_R) \tag{1.3}$$

where D_μ is the covariant derivative describing the electroweak interactions between the gauge fields and the fermion fields in which g and g' are the coupling constant of the $SU(2)_L$ and $U(1)_Y$ groups respectively.

$$D_\mu = \partial_\mu + \imath g T^a W_\mu^a + \imath g' \frac{Y}{2} B_\mu \tag{1.4}$$

The first two terms in Equation 1.3 are the kinetic energy terms of W_μ^a and B_μ generators with

$$W_{\mu\nu}^a = \partial_\mu W_\nu^a - \partial_\nu W_\mu^a + g \varepsilon_{jk}^a W_\mu^j W_\nu^k \tag{1.5}$$

$$B^{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu \tag{1.6}$$

where ε^a_{jk} is the antisymmetric Levi-Civita tensor. The bi-linear term $g\varepsilon^a_{jk}W^j_\mu W^k_\nu$ shows the self interaction of the W^a_μ gauge bosons. The *non-Abelian* $SU(2)_L$ group allows the coupling of W^a_μ fields with itself whereas the *Abelian* nature of $U(1)_Y$ group prohibits the self interaction of B_μ. Linear combination of the W^a_μ and B_μ bosons are realized with the physical $W^\pm$, Z^0 and γ bosons as follows:

$$W^{\pm,\mu} = \frac{1}{\sqrt{2}}(W^\mu_1 \pm \imath W^\mu_2) \tag{1.7}$$

$$\gamma^\mu = \sin\theta_W W^\mu_3 + \cos\theta_W B^\mu \tag{1.8}$$

$$Z^{0,\mu} = \cos\theta_W W^\mu_3 - \sin\theta_W B^\mu \tag{1.9}$$

where θ_W, the weak mixing angle also known as the Weinberg angle, is a measure for the mixing between the W_μ and B_μ terms.

All the fermions and bosons introduced so far are massless. Unfortunately, adding a Dirac mass term would violate the gauge invariance of these fields. The masses of the weak gauge bosons are generated by introducing an additional term in Equation 1.3 associated to a complex scalar field through the Brout-Englert-Higgs [43,44] mechanism. This scalar field, called the Higgs field, causes the Spontaneous Symmetry-Breaking (SSB) of the $SU(2)_L \times U(1)_Y$ vacuum symmetry to $U(1)_{EM}$, which is the gauge symmetry of electromagnetism. The Lagrangian for this field is

$$\mathcal{L}_H = (D^\mu\phi)^\dagger(D_\mu\phi) - V(\phi) \tag{1.10}$$

which consists of a kinetic term and a Higgs potential $V(\phi)$, with

$$V(\phi) = -\mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{1.11}$$

The first term in Equation 1.11 is associated to the mass of the field, while the second term describes the self-interaction of the scalar field. As a result λ must be non-zero to maintain the stability of the vacuum. If $\mu^2 > 0$, the potential has only the trivial minimum $\phi = 0$. For the case $\mu^2 < 0$, the field has a non-zero vacuum expectation value, ν, leading to non-trivial minima $\phi = \frac{\sqrt{\nu}}{2}$ which spontaneously breaks the $SU(2)_L \times U(1)_Y$ symmetry. After SSB, three of the four degrees of freedom of Higgs field are absorbed by the massless $W^\pm$ and Z^0 bosons which consequently acquire mass. The remaining degree of freedom constitutes the Higgs boson, a massive scalar particle. The electromagnetic gauge boson remains massless. The calculation as demonstrated in reference [45] results in the mass terms

$$m_W = \frac{1}{2}\nu g, \qquad m_Z = \frac{1}{2}\nu\sqrt{g^2 + g'^2}, \qquad m_\gamma = 0, \qquad m_H = \sqrt{2\lambda\nu^2} \tag{1.12}$$

The masses of the fermions are generated by introducing a Yukawa coupling [46] that describes the interaction between the Higgs and fermion fields. This term can be expressed as

$$\mathcal{L}_{Yukawa} = \sum_{fermions} \lambda_f(\bar{\psi}_L \phi \psi_R + \bar{\psi}_R \phi \psi_R) \tag{1.13}$$

The mass of the fermions are expressed as

$$m_f = \lambda_f \frac{v}{\sqrt{2}} \tag{1.14}$$

The fermion with highest coupling to the Higgs boson is the top quark.

1.1.3. Strong Interaction

The strong interaction is based on the $SU(3)_C$ symmetry that describes the strong interaction between coloured quarks and gluons. The existence of the Δ^{++} particle was observed [47], which is a three u-quarks bound state i.e. three quarks of the same flavour and spin. This particle seemed to break the Pauli principle which states that two or more identical fermions cannot have the same set of quantum numbers. The additional degree of freedom introduced by the colour charge therefore defines antisymmetric states formed by three quarks with same flavour and spin. The QCD Lagrangian is given as

$$\mathcal{L}_{QCD} = -\frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu} + \bar{\psi}(\imath\gamma^\mu D_\mu - m)\psi \tag{1.15}$$

where ψ is the fermion field with mass m. Here, a runs from 1 to 8 representing eight generators of $SU(3)_C$ symmetry group, rather than 1 to 3 as in the $SU(2)_L$ case. D_μ is the covariant derivative, introduced to maintain the gauge invariance of the Lagrangian.

$$D_\mu = \partial_\mu - \imath g T^a A^a_\mu \tag{1.16}$$

$$F_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu - \imath g f^{abc} A^b_\mu, A^c_\nu \tag{1.17}$$

where A^a_μ are the gauge fields and g is the gauge coupling parameter. f^{abc} are the so-called structure constants of the group, defined from the commutation relation of the Gell-Mann matrices [48]. The last term in Equation 1.17 represents the self-coupling of the gluons arising from the non-Abelian nature of $SU(3)_C$. The self interaction of gluons results in a more complicated QCD theory as compared to the QED where photons being electrically neutral do not interact with each other.

The self-coupling of gluons is the reason behind one of the important characteristics of QCD, known as *colour confinement*. It implies that the quarks are always observed to be confined

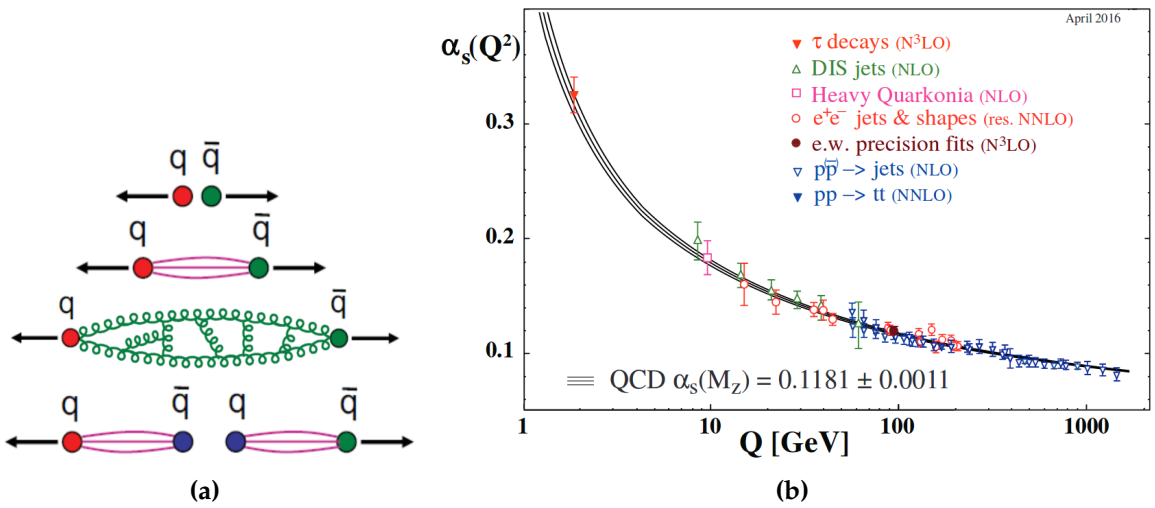


Figure 1.2.: Pictorial presentation of effect of gluon-gluon interaction forming colour flux tube resulting in colour confinement in QCD [50] (left) and running of the strong coupling constant with respect to the energy scale (Q) of the interaction [51] (right).

in a colourless bound state and never as a free coloured state. If enough energy is applied to separate a bound state of say $q\bar{q}$ state, as the distance between the quarks increases the interaction also becomes very intense. If they are about to separate, a colour flux tube of the interacting gluons is formed as demonstrated in Figure 1.2a. The energy in this tube is linearly proportional to the distance between the quarks, hence to separate them infinite energy would be needed. The energy contained in the tube is sufficient to create a light quarks and antiquark pair. The new final state consists of two colour-neutral bound states.

The strength of strong interaction is determined by the strong coupling constant α_S, which is the fundamental parameter of QCD. Similar to the EM coupling constant α, the dependence of α_S on the momentum transfer Q^2 (Figure 1.2b) makes it a running coupling constant but its behaviour is very different from $\alpha(Q^2)$. At very high energy which corresponds to smaller distance and high momentum transfer, the strong coupling constant becomes small and decreases asymptotically. This feature of QCD is called *asymptotic freedom* i.e. in the $Q^2 \to \infty$ limit, the coupling $\alpha_S \to 0$, which means the quarks behave as quasi-free particles. The calculations in QCD are expanded in orders of α_S as discussed further in Section 1.2.1. Therefore, only at high Q^2 when the value of α_S is very small, the perturbation theory can be applied to perform the calculations. Whereas, at energy scales of $\Lambda_{QCD} \approx 200\,\text{MeV}$ [49] and lower, the coupling reaches unity and perturbative QCD (pQCD) is not applicable.

1.2. Phenomenology of Proton-Proton Collision

A proton is not an elementary particle, rather a composition of two up quarks and one down quark. The quantum numbers associated to a proton are determined by these three valence quarks (*uud*). It also consists of a sea of gluons being exchanged by the valence quarks and sea of quark-antiquark pairs produced through gluon interactions. The valence quarks and sea of gluons and quarks are called *partons*.

In the high energy proton-proton collision, when the protons are accelerated at high energies, these constituent partons attain fraction of the proton momentum and participate in the collision. From the previous concept of asymptotic freedom, the pp collision can be divided into high momentum transfer (*hard scattering* process) and low momentum transfer (*soft* interactions). The short distance hard scattering process is described by the perturbative QCD theory while the soft interaction that occur between the remnants after the hard scattering and further among the newly created quarks and gluons, are subject to non-perturbative QCD. Figure 1.3 shows a typical pp collision event with all the ongoing processes as explained in the figure caption and in details below.

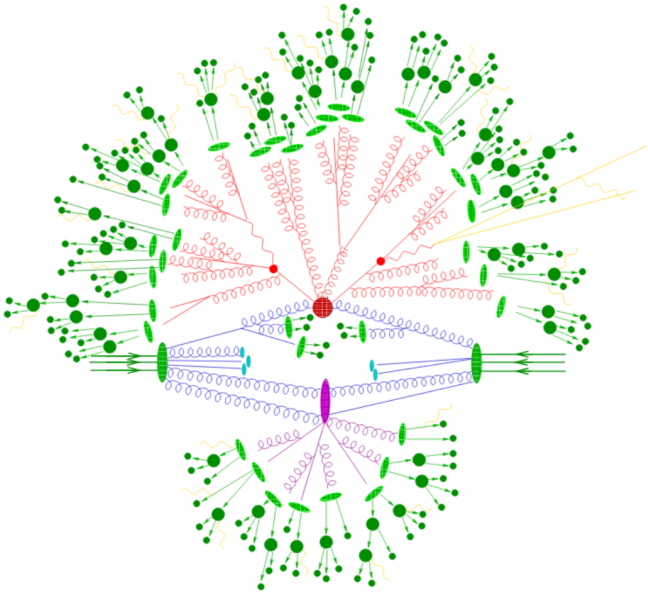


Figure 1.3.: Diagrammatic presentation of $t\bar{t}$ production in a pp collision event taken from Reference [52] and edited. Two incoming protons are shown with dark green ellipses. The big red blob represents the hard scattering followed by the decay of the top quarks (small red blob). Initial state radiation can be seen just below the big red blob in blue colour and final state radiation in red colour. The light green blobs show the hadronization stage and the dark green circles correspond to hadron decays. The purple ellipse represents a secondary interaction between low momentum partons and cyan colour blobs indicate the hadronization of beam remnants. Photon radiation occurs at any stage (yellow). Each term is mentioned in the following text.

1.2.1. Partonic Cross-Section

The foremost need for a hadron collider is the QCD calculation to relate the incoming initial state to the outgoing final state. This is achieved by the scattering matrix S_{fi} which describes the transition between the initial state i and the final state f based on the physical information of the states, quantum numbers, coupling and spin of the particles involved. The matrix element (ME) that represents the transition probability amplitude from the initial state ψ_i to a final state ψ_f is

$$\mathcal{M} = \int d\phi_n \psi_f^\dagger S_{fi} \psi_i \tag{1.18}$$

where $d\phi_n$ is the n-particle final state phase space. The *partonic cross-section* of the hard scattering, which is the probability for the production of a process in pp collision $\hat{\sigma}$, is then given by

$$\hat{\sigma} = \int \frac{1}{F} |\mathcal{M}|^2 d\phi_n \tag{1.19}$$

where F is the incoming particle flux. Since the partons which participate in the collision acquire only a fraction of the initial momentum of the proton, extra factors need to be introduced in the inclusive cross-section calculation which will be explained later in the text. The inclusive cross-section for the process of interest in a pp collision (say, $pp \to X$) is defined in terms of partonic cross-section $\hat{\sigma}_{ab\to X}$ as

$$\sigma_{pp\to X} = \sum_{a,b} \int dx_a \, dx_b \, f_a(x_a, \mu_F^2) \, f_b(x_b, \mu_F^2) \, \hat{\sigma}_{ab\to X}(x_a, x_b, \alpha_S(\mu_R^2), \mu_R, \mu_F) \tag{1.20}$$

where the sum runs over all the partons a and b which can initiate the process. x_a and x_b are the fraction of the momentum of the incoming colliding protons, carried by the partons a and b. The function $f_{a(b)}$ called the Parton Distribution Function (PDF) is the probability of the parton $a(b)$ to carry the momentum fraction $x_a(x_b)$ for an energy scale Q.

The ME for the transition has perturbative expansion in α_S. The partonic cross-section $\hat{\sigma}_{ab\to X}$ in Equation 1.20 can be expanded in fixed order series in α_S as

$$\hat{\sigma}_{ab\to X} = [\, \hat{\sigma}^{(0)} + \alpha_S(\mu_R^2)\hat{\sigma}^{(1)} + ... + \alpha_s^N(\mu_R^2)\hat{\sigma}^{(N)} \,] \tag{1.21}$$

where $\hat{\sigma}^{(0)}$ is the partonic cross-section at leading order (LO), $\hat{\sigma}^{(1)}$ at next-to-leading order (NLO) and so on which implies $\hat{\sigma}^{(N)}$ is the contribution at specific order N in the perturbation theory. Therefore, each term would correspond to a set of unique processes contributing to the total interaction with a fixed power of α_S and can be represented by a set of Feynman diagrams

as in Figure 1.4. Figure 1.4a and 1.4b shows the LO term and one NLO term of Equation 1.21. Adding higher order terms increases the precision but would also introduce all diagrams up to that order which complicates the calculations. So the measurements are performed at certain finite terms like LO, NLO or NNLO.

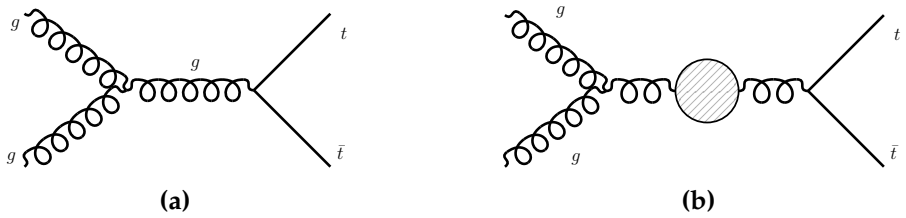


Figure 1.4.: Example of Feynman diagrams for $t\bar{t}$ production (a) at born level i.e. LO and (b) for virtual correction.

Processes involving self-energy loop corrections (Figure 1.4b) possibly lead to the divergence of the integrals of internal loops momenta. This is called ultraviolet divergences and is dealt with by the renormalisation procedure. First a cut-off parameter is introduced to split the integral into finite and infinite terms using regularization. Then the divergent term is absorbed by adding counter terms in the renormalisation step via redefinition of particle mass, coupling constant [24]. This leads to the introduction of an arbitrary scale known as the *renormalisation scale* (μ_R). As a consequence, the coupling constant α_S depends on the choice of μ_R and is characterized by its *running* behaviour as discussed previously in Section 1.1.3.

As stated earlier the perturbation theory is applicable only at high energy. The evolution of pp collision also consists of long distance soft interactions where perturbative QCD cannot be applied. These short and long distance effects are separated by an arbitrarily chosen energy scale referred to as the *factorisation scale* (μ_F) using the factorisation theorem [53].

As the energy scales can be chosen arbitrarily, these are often fixed based on the hard interaction of the process of interest such as invariant mass of the final state particles and are usually set to be $\mu_R = \mu_F$. The dependence on the choice of these energy scales introduces the so called scale uncertainties in the cross-section calculations.

Parton Distribution Function

To calculate the *cross-section*, the knowledge of momentum fraction carried by a parton is required. The probability of a parton a to carry the momentum fraction x_a of the proton at an energy scale Q^2 is parametrized as Parton Distribution Functions (PDFs). The evolution equations for the PDFs at certain Q^2 scales are empirically calculable using the DGLAP

equations [54, 55].

$$\frac{\partial f_a(x,Q^2)}{\partial \log Q^2} = \frac{\alpha_S}{2\pi}\int_x^1 \frac{dz}{z}\left[P_{ab}(z,\alpha_S)b\left(\frac{x}{z},Q^2\right)\right] \tag{1.22}$$

where $f_a(x,Q^2)$ is either a quark or gluon PDF and $P_{ab}(z,\alpha_S)$ denotes the splitting function [56] that can be expanded in powers of α_S. However, these equations cannot predict the x dependence of the PDF at a given Q^2 and is therefore determined from experimental data from lepton-hadron collider ZEUS [57] and H1 [58] at HERA accelerator, hadron collider D∅ [59] CDF [60] at Tevatron and also from LHC. PDF sets are then obtained by fitting on a significantly large dataset from these experiments. Various collaborations work for the improvement of PDF by fitting to most recent data. Since the assumptions and limits used for the DGLAP equation by different groups are slightly different, an associated theoretical uncertainty is also estimated. Figure 1.5 shows the dependence of PDF on x at two different energy scale provided by the NNPDF [61] group.

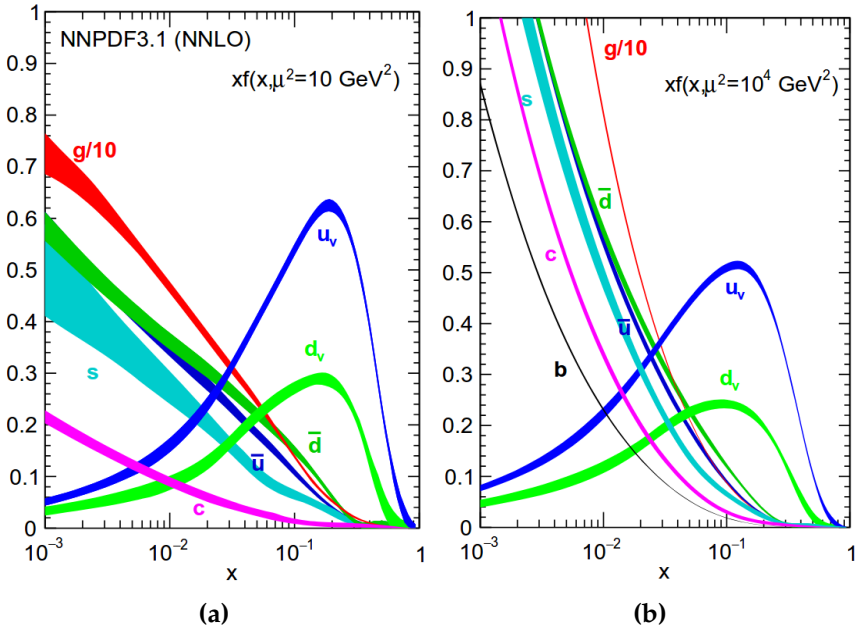


Figure 1.5.: Parton distribution function for proton at NNLO in α_s^2, from the NNPDF group, evaluated at $Q^2 = 10\,\mathrm{GeV}^2$ (left) and $Q^2 = 10^4\,\mathrm{GeV}^2$ (right) [61].

1.2.2. Parton Shower

The hard scattering is evaluated from the partonic cross-section at a fixed order in pQCD with a certain choice of the renormalisation and factorisation scales and a given PDF set. The partons produced in the hard process start moving away which increases the colour field strength emitting gluons and quark-antiquark pairs. The resulting partons can successively

split as $g \to gg, g \to q\bar{q}$ or $q(\bar{q}) \to q(\bar{q})g$ and results in a shower or cascade of partons which is called Parton Shower (PS). The shower can originate from both the initial and the final state of the hard process, these are called the *initial state radiation* (ISR) and *final state radiation* (FSR) respectively.

The evolution of the cascade of partons can be determined by the implication of Sudakov form factor [62] which describes the probability of a parton to evolve from scale q_1 to q_2 without splitting. The differential probability $dP_a(z, q^2)$ for a parton a to split into two partons at an energy scale q^2 is given by

$$dP_a(z, q^2) = \frac{\alpha_S}{2\pi} \frac{dq^2}{q^2} dz \hat{P}_{ba}(z) \tag{1.23}$$

where the produced parton b carries the momentum fraction z of the parton a. By knowing the probability of emission and no emission, the PS cascade can be formed [63] using the DGLAP evolution as in Equation 1.22. Figure 1.6 shows the evolution of a parton from scale q_0 to q' without splitting, then a splitting at q' and again from q' to q without splitting. This is also referred to as *fragmentation*.

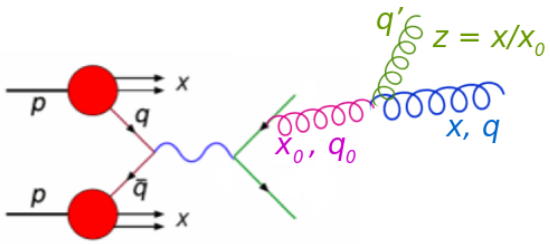


Figure 1.6.: Pictorial representation of the evolution of parton shower in pp collision.

The PS starts at an energy scale of hard scattering and each emission gradually lowers down the energy of the event to a soft and collinear emission. The mathematical calculation [63] results in diverging logarithmic terms $\left(\frac{\alpha_S}{2\pi} \log \frac{t}{t_0}\right)^n$. Therefore, the calculation includes a few terms such as leading-log (LL), next-to-leading-log (NLL) and so on but it is only valid for small-angle gluon emissions.

1.2.3. Hadronization

At lower energy scale of $\mathcal{O}$(1-2 GeV) [64], the shower confines into hadrons because of colour confinement. The partons start to hadronize and form colour-neutral baryons and mesons. The excited hadrons further decay into many lower energy states and lead to generation of observed final state hadrons, this process is referred to as *hadronization*. The hadronization

process is non-perturbative and is described by phenomenological models. As mentioned, the energetic quarks can produce a cascade of $q\bar{q}$ pairs followed by subsequent additional soft radiations, hadronization and decay, leading to a narrow cone of hadrons referred to as *jets*.

1.2.4. Underlying Events

The explanations so far have not considered any parton interactions which did not participate in the hard scattering. The Underlying Event (UE) is everything except the hard scattering process such as the proton remnants and multi-parton interaction. The UE mostly populate the low energy non-perturbative QCD but sometimes an energetic multi-parton interaction gives rise to a high momentum particle as well in perturbative regime. Hence it is described by phenomenological models with free parameters which are tuned based on experimental data.

In conclusion, the simulation of a physics processes in the pp collision takes into account all the effects mentioned in Section 1.2. The combination of the hard scattering process with the soft interaction describes the full phenomena. This combination is not trivial because of the presence of same configuration/diagrams from PS and higher order ME. These double counting and interference are taken care in simulation using a method called *matching and merging*, explained in Chapter 3.

1.3. The Top Quark

The top quark, discovered in 1995 by the CDF [65] and D∅ [66] experiments at the Tevatron, is the heaviest of all the known elementary particles ($m_t = 173.0 \pm 0.4\,\text{GeV}$) [32]. Its large mass results in several unique properties of the top quark. As is deduced from Equation 1.14, it has strong coupling to the Higgs boson which is close to 1.

$$\lambda_t = \sqrt{2}\frac{m_t}{v} \approx 1 \tag{1.24}$$

The top quark has very short lifetime ($\tau_t \sim 5 \times 10^{-25}$ s) [32] so it decays before hadronization, unlike other quarks which are confined in bound sates of hadrons. This property gives an opportunity to study a *bare* quark.

1.3.1. Top-Quark Pair Production

LHC is sometimes also referred to as the "top-factory" because of large cross-section of top quark pair production. Top quark pairs are produced through strong interaction via two

mechanisms: gluon-gluon (gg) fusion and quark-antiquark ($q\bar{q}$) annihilation. Figure 1.7 shows the LO Feynman diagrams of the $t\bar{t}$ production.

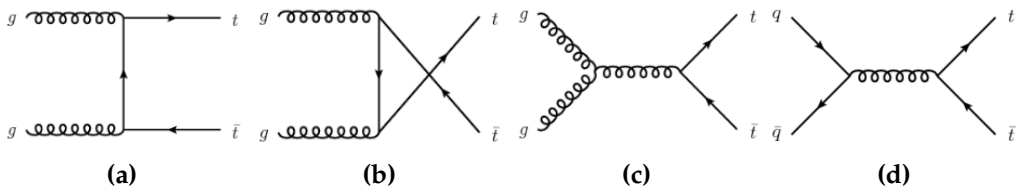


Figure 1.7.: LO Feynman diagrams for the top quark production through the (a), (b), (c) gluon-gluon fusion and (d) quark-antiquark annihilation.

At the LHC with center-of-mass energy ($\sqrt{s}$) of 13 TeV, around 85-90% of the top quark pairs are produced through gg fusion whereas the remainder stems from $q\bar{q}$ annihilation. However, the dominant production of $t\bar{t}$ at the Tevatron ($p\bar{p}$ collider) was through $q\bar{q}$ annihilation. This mechanism is favoured at Tevatron because of collision occurring between the valence quarks and antiquark coming from the proton and anti-proton.

Inclusive and Differential Cross-sections

As inferred from Equation 1.20, the inclusive $pp \to t\bar{t}$ cross-section depends on the mass of the top quark and the square of the center-of-mass energy of the collision, s. The theoretical computation is performed at NNLO in QCD including resummation of next-to-next-to leading logarithmic (NNLL) soft gluon terms with top++2.0 [67–70]. The predicted central value at 13 TeV and $m_t = 172.5$ GeV [71] is

$$\sigma_{t\bar{t}} = 831.76^{+19.77}_{-29.20}\,(\text{scale unc.}) \pm 35.06\,(\text{PDF} + \alpha_S\,\text{unc.})\,\text{pb} \tag{1.25}$$

The quoted scale uncertainties arise from the variations of renormalisation and factorisation scales while the other comes from the PDF and strong coupling constant. Figure 1.8 shows the theoretical as well as experimental measurements of the $t\bar{t}$ cross-section at different center-of-mass energies for the LHC and Tevatron colliders.
In addition to the inclusive cross-section, measurement of the *differential cross-section* of $t\bar{t}$ provides a stringent test of the pQCD. It is the measurement of cross-section as a function of some kinematic observable like the momentum, angle or invariant mass of $t\bar{t}$ system. If x is a kinematic quantity, the differential cross-section can be given as $\frac{d\sigma_{t\bar{t}}}{dx}$. These distributions help in analysing the observable in divided phase spaces (bins). It provides test of the QCD prediction to further constrain the SM and to identify possible indication of new physics that could hide in limited phase space without any pronounced deviation in total $\sigma_{t\bar{t}}$.

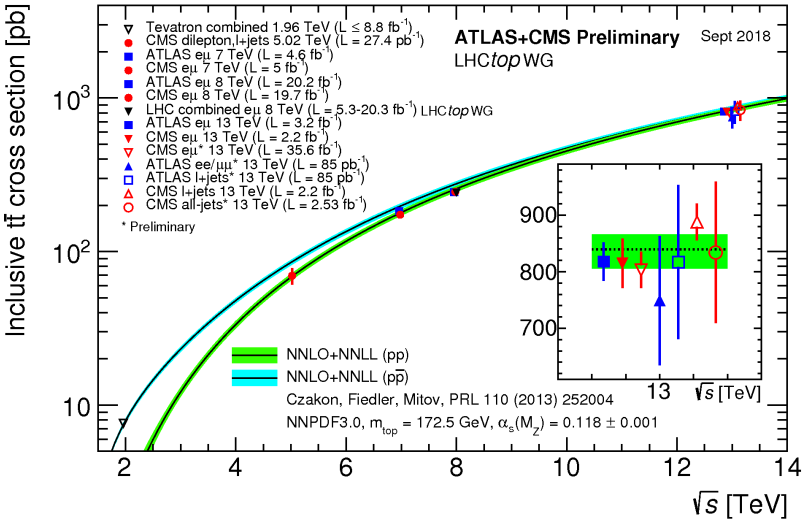


Figure 1.8.: Summary of LHC and Tevatron measurements of the top-pair production cross-section as a function of the centre-of-mass energy compared to the NNLO QCD calculation complemented with NNLL resummation (top++2.0). The theory band represents uncertainties due to renormalisation and factorisation scale, parton density functions and the strong coupling. The measurements and the theory calculation are quoted at $m_t = 172.5\,\text{GeV}$ [71].

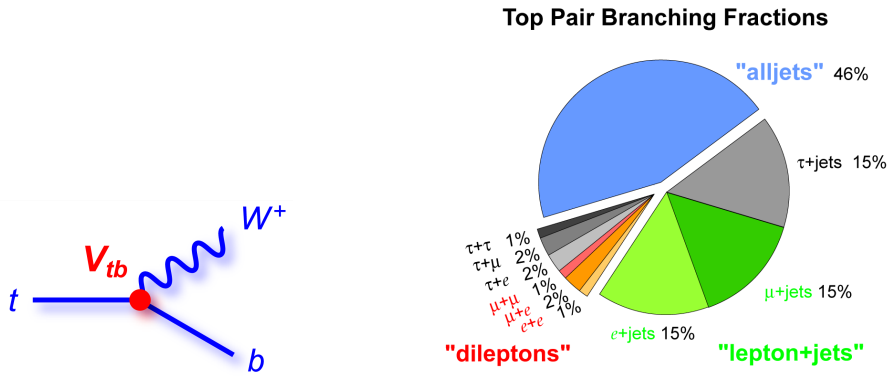
1.3.2. Decay of Top-Quarks

The top quark decays through electroweak interaction, decaying almost always into a W boson and a b-quark. The probability of a top quark to decay into $W^{\pm}$ and a down-type quark of any generation is obtained using the Cabibbo-Kobayashi-Maskawa (CKM) matrix [72, 73]. The 3×3 unitary CKM matrix describes the mixing of quarks by obtaining mass eigen-state of the quarks from the corresponding weak eigen-state. The elements of the CKM matrix are fundamental parameters of the SM and are extracted from a global fit to the available measurements.

$$V_{\text{CKM}} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \simeq \begin{pmatrix} 0.9745 & 0.2245 & 0.0037 \\ 0.2244 & 0.9736 & 0.0421 \\ 0.0089 & 0.0413 & 0.9991 \end{pmatrix} \tag{1.26}$$

The values in Equation 1.26 are taken from [32] but the uncertainty on the measurement is not quoted here. The elements of the matrix determine the strength of the flavour changing weak decays of each quark. The numerical values of the parameters $|V_{td}|$ and $|V_{ts}|$ are way smaller than $|V_{tb}|$ which is almost equal to 1. Thus the top decays $\sim 100\%$ of the time into a b-quark and a W boson (Figure 1.9a).

The final state of a top quark decay is therefore dependent on the decay mode of the W boson. The W boson decays either leptonically into a charged lepton and the corresponding neutrino or hadronically into a pair of quarks ($q\bar{q}'$ where $q = u, c$ and $\bar{q}' = \bar{d}, \bar{s}, \bar{b}$). Since the $t\bar{t}$ final



(a) **(b)**

Figure 1.9.: (a) Pictorial representation of the top quark decay. (b) Pie chart representation of $t\bar{t}$ decay in different channels with the branching ratios.

state contains two W bosons, the decay modes are classified as following and the branching ratio[2] ($\mathcal{BR}$) is shown in Figure 1.9b.

- *fully-hadronic* channel referes to the decay of both the W bosons into $q\bar{q}'$ pairs (Figure 1.10a) which corresponds to a $\mathcal{BR}$ of $\sim$46%.
- *semi-leptonic* or *leptons+jets* channel is where one W decays to a $q\bar{q}'$ pair while the other into a lepton and neutrino with a $\mathcal{BR}$ $\sim$45%.
- *di-leptonic* channel is the one where both the W bosons decay into leptons (e, μ or τ) with their associated neutrino (ν_e, ν_μ or ν_τ). The total branching ratio of this final state is $\sim$9%. The final state with one W decaying to an electron and another into a muon with their corresponding neutrinos is referred to as $e\mu$ channel and corresponds to only $\sim$2% $\mathcal{BR}$.

Figure 1.10 represents the three possible final states of $t\bar{t}$ decays. This thesis concerns the $e\mu$ channel.

1.4. $t\bar{t}$ with Additional b-jets

In this thesis, the cross-section of top quark pair production in association with additional b-jets which is referred to as $t\bar{t}+b$-jets is measured. The b-jets produced in addition to the two b-jets coming from the top quark decays are called additional b-jets and can arise from various sources. Major production is from QCD radiation with gluons splitting into $b\bar{b}$ pair in the pp collision (Figure 1.11a). Other processes with the same final state are: the production of a Higgs boson in association with top quark pair where the Higgs boson subsequently decays

[2]The branching ratio determines how likely a particle decays in a certain decay mode. Clearly, the value lies between 0 to 1 but is often quoted in percentage also.

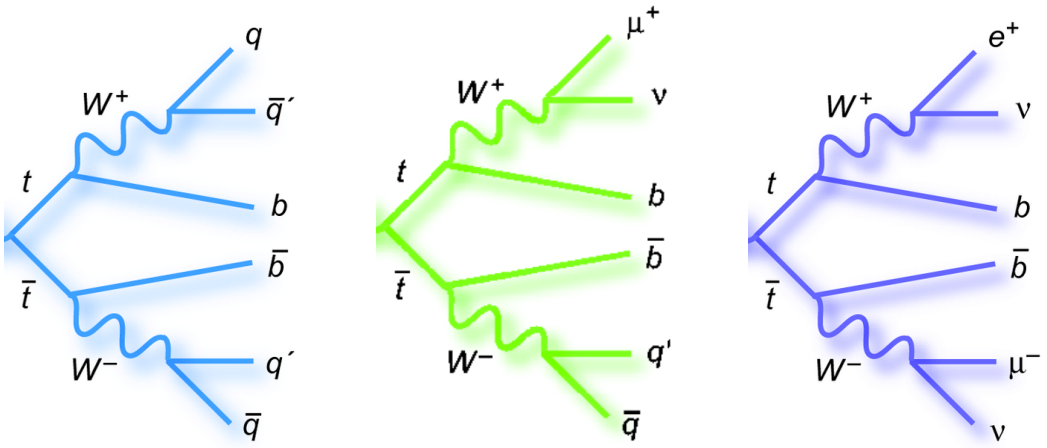


Figure 1.10.: Feynman diagrams for decay modes of the top quark pair in (a) fully hadronic, (b) semi-leptonic and (c) dileptonic channels.

into a b-quark pair $t\bar{t}H(H \to b\bar{b})$ (Figure 1.11b) and similarly the production of a Z boson in association with $t\bar{t}$ and subsequent decay as $t\bar{t}Z(Z \to b\bar{b})$ (Figure 1.11c). The production of $t\bar{t}W$ where W decays into $q\bar{q}'$ (Figure 1.11d) with one of the quarks as a b-quark has very small cross-section and does not actually contribute to the $t\bar{t} + b$-jets analysis but is mentioned for completion.

***$t\bar{t}H(b\bar{b})$*:** Since the discovery of Higgs boson, the measurement of its interactions permits to

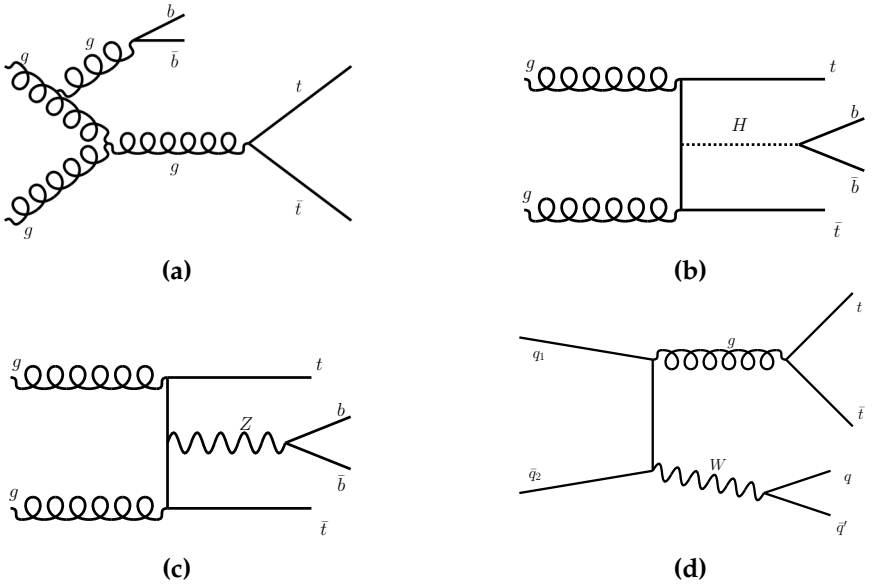


Figure 1.11.: Example of Feynman diagrams leading to $t\bar{t}b\bar{b}$ final state (a) QCD $t\bar{t}b\bar{b}$ production, (b) $t\bar{t}H(b\bar{b})$, (c) $t\bar{t}Z$ $(Z \to b\bar{b})$ and (d) $t\bar{t}W$ $(W \to q\bar{q}')$.

probe the spontaneous symmetry breaking mechanism. The production of Higgs boson in association with $t\bar{t}$ allows direct measurement of the top quark Yukawa coupling. The $\mathcal{BR}$ of $H \to b\bar{b}$ is the highest which corresponds to 58.1% [64]. Hence, it provides the best statistical precision for the direct top quark Yukawa coupling measurement, thereby rendering essential test to the mass-generation mechanism in the heavy quark sector. However, the predicted cross-section for $t\bar{t}H$ process at 13 TeV is $507^{+81.6}_{-98.9}$ fb [64] calculated for $m_H = 125.09$ GeV at NLO QCD including electroweak correction. Although the $t\bar{t}H(b\bar{b})$ process has the largest $\mathcal{BR}$, it has roughly 1600 times smaller cross-section than the $t\bar{t}$ process. Consequently, $t\bar{t}H(b\bar{b})$ suffers from the dominant and almost irreducible background from QCD induced $t\bar{t}b\bar{b}$ process. Therefore it is very important to properly estimate the QCD $t\bar{t}b\bar{b}$ contribution.

$t\bar{t}Z(b\bar{b})$: The predicted NLO cross-section with electroweak correction for $t\bar{t}Z$ production is 839^{+87}_{-99} fb [64]. However, the $\mathcal{BR}$ of $Z \to b\bar{b}$ is only 15.1% [32] which results in a very small contribution of $t\bar{t}Z(b\bar{b})$ to the $t\bar{t} + b$-jets process.

$t\bar{t}W(q\bar{q}')$: The cross-section for $t\bar{t}W$ is also very small as compared to the $t\bar{t}$ process corresponding to 601^{+58}_{-53} fb as predicted at NLO with electroweak correction [64]. Even though the W boson decays 67.4% [32] of the time into hadronic channels, the decay of W into a b-quark final state is negligible. The dominant decay modes are CKM favoured (Equation 1.26) $c\bar{s}$ and $u\bar{d}$ because of high values of $|V_{ud}|$ and $|V_{cs}| \simeq 0.97$. Therefore the cross-section for $t\bar{t}W$ where W decays further into b-quarks has negligible contribution in $t\bar{t} + b$-jets.

The production of $t\bar{t}$ with additional b-jets provides an important test of the QCD predictions. One example is the challenge to calculate the amplitude for the process in Figure 1.11a due to non-negligible mass of the b-quark which implies two scales in the QCD calculation. Progress in the theory prediction for $t\bar{t}b\bar{b}$ production is the key prerequisite to improve the sensitivity for $t\bar{t}H(b\bar{b})$ process. The final state of $t\bar{t}b\bar{b}$ consists of two charged leptons, the corresponding two neutrinos, two b-quarks from the top quarks decays and at least one additional b-quark which makes its study a challenging task [74]:

- **Complex process:** The final state consists of 8 particles, which implies the existence of many partonic channels. The calculation of the full process is intractable in many computations (MC generators; mentioned in Chapter 3), they are able to compute double resonant diagrams only where both the top quarks are on-shell and considered stable with $\mathcal{O}(\alpha_S^4\alpha^4)$ as shown in Figure 1.12.
- **Sensitive to scale choice:** Therefore, the contribution usually generated by MC is proportional to α_S^4. The value of α_S depends on the renormalisation scale as discussed in Section 1.2.1, hence the calculation is affected severely by the choice of μ_R.

- **Contain multiple scales:** The final state consists of two strong couplings one with the top quark (red dots) and another with bottom quark (yellow dot) as the mass of the b-quark is also non-negligible. The presence of multiple scales spoils the convergence of the perturbative expansion [75].

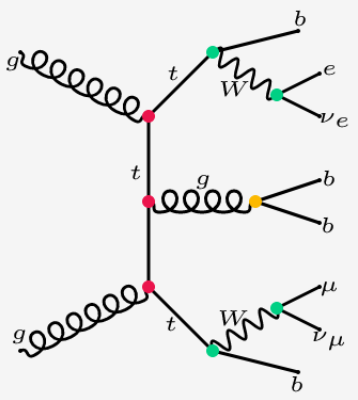


Figure 1.12.: Leading order diagram for the $pp \to t\bar{t}b\bar{b} \to W^+bW^-b\bar{b}\bar{b} \to e\nu_e\mu\nu_\mu b\bar{b}b\bar{b}$ in the SM showing the complexity of the process containing 8 particles in final state. The red and yellow dots represent the strong coupling (α_S) whereas green dots show the electroweak couplings (α).

To deal with all these complications, different approximations and calculation schemes are employed in the MC generators which are discussed in Chapter 3.

Chapter 2.

The ATLAS Experiment at the LHC

The analysis performed in this thesis relies on the proton-proton (pp) collision data acquired by the ATLAS experiment at the Large Hadron Collider (LHC). This chapter provides an introduction to the experimental setup. Section 2.1 presents an overview of the LHC accelerator used to accelerate the protons to very high energy and collide them to generate the data. Section 2.2 explains the ATLAS detector and its subsystems, placed at one of the four collision points of the LHC ring to collect the produced data. Section 2.3 describes the conditions of pp beams collided at the ATLAS detector in 2015 and 2016.

2.1. The Large Hadron Collider

The Large Hadron Collider (LHC) [76] situated at the European Organization for Nuclear Research (CERN) is the world's largest and most complex particle accelerator residing in a 27 km circular tunnel at the France-Switzerland border. It is designed to accelerate two beams of protons and also lead (Pb) ions (not used in this thesis) in opposite direction in two parallel rings. The protons are injected in both beam-pipes in bunches which contain around 10^{11} protons each. The bunches are spaced by 25 ns. They are guided around the accelerator ring by a strong magnetic field maintained by superconducting electromagnets.

Before getting injected into the LHC ring, the proton beams undergo a series of pre-accelerations. The first accelerator is the linear accelerator 2 (Linac2) which accelerates the protons to 50 MeV. The protons are then accelerated to 1.4 GeV by the Proton Synchrotron Booster (PSB) and fed to the Proton Synchrotron (PS) where these are accelerated to 25 GeV. They are further accelerated to 450 GeV in the Super Proton Synchrotron (SPS) and then transferred to the LHC beam-pipes. Each of the beams is finally accelerated to an energy of 6.5 TeV (for Run-2) and are collided, which implies the centre-of-mass collision energy, $\sqrt{s} = 13$ TeV. The collision points are the positions of four major detectors across the accelerator ring, namely: ALICE [77],

ATLAS [78], CMS [79] and LHCb [80]. A schematic diagram of the complex LHC machine with its pre-accelerators and detectors is shown in Figure 2.1.

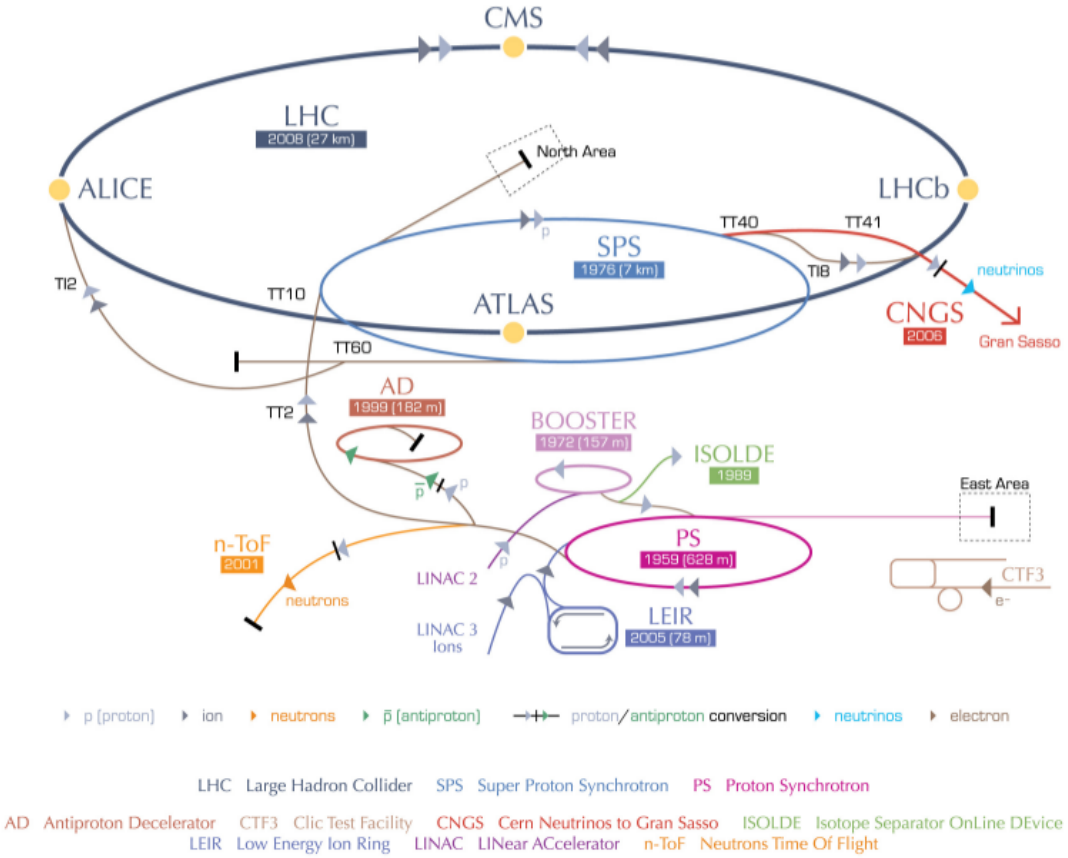


Figure 2.1.: A schematic diagram of the LHC machine with its pre-accelerators and major detectors [81].

2.2. The ATLAS Detector

With a dimension of $46 \times 25 \times 25$ m and weight 7,000 tonnes, ATLAS (A Toroidal LHC ApparatuS) [78] is the largest volume particle detector ever constructed. It is one of the two general purpose detectors (the other being the CMS detector) built to study the pp collisions at the LHC. It is situated 100 m below the ground near the main CERN site. The collision of proton beams occurs at the centre of the detector producing a huge number of particles which fly out in all possible directions. To cover this, the detector is designed with cylindrical and forward-backward symmetric geometry around the interaction point. The cylindrical layers around the beam-pipe are referred to as barrels. To improve the coverage in forward direction, end-caps are placed perpendicular to the beam-pipe at both ends of the detector. As the out-going particles posses different properties, these layers are made up of different

interaction materials and techniques to detect and identify them. The sub-detectors of ATLAS - the inner detector, the calorimeters and muon spectrometer are illustrated in Figure 2.2 and explained in the following sections.

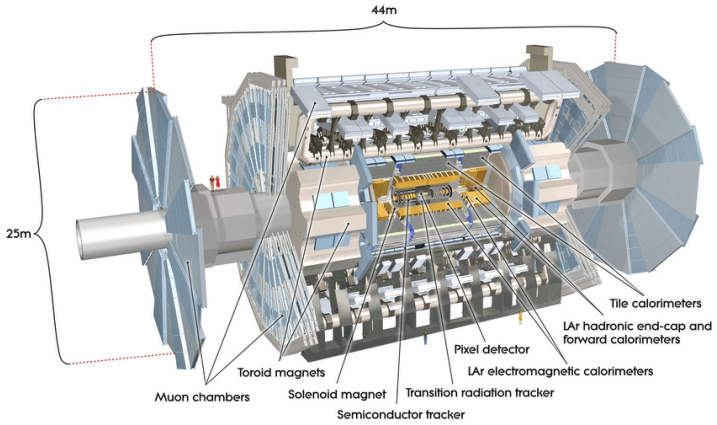


Figure 2.2.: Layout of the ATLAS detector and its main components [78].

2.2.1. Coordinate System

The ATLAS experiment uses a right-handed Cartesian coordinate system where the nominal interaction point at the centre of the detector is defined as the origin, the beam direction is considered as the z-axis and transverse to the beam axis is the x-y plane. The positive x-axis points towards the centre of the LHC ring and the positive y-axis points vertically upwards. The cylindrical geometry of the detector also facilitates to work with polar coordinates (r, θ, ϕ). The angle ϕ[1] is the azimuthal angle measured around the beam axis, while θ[2] is the polar angle measured from the beam axis.

Since the proton beams are accelerated only in the z-direction and collided head-on, the initial momentum of the system in the x-y plane is negligible. By the law of conservation of momentum, the vector sum of momenta of all final state particles in the transverse plane must be equal to zero. Therefore, observables such as transverse momentum that are Lorentz invariant to boosts along the z-axis are relevant to measure.

The momentum of a particle is expressed as:

$$\boldsymbol{p} = (p_x, p_y, p_z) \tag{2.1}$$

[1]The angle ϕ is measured around the beam axis counter clock-wise from the positive x-axis in the x-y plane
[2]The angle θ is measured from the positive z-axis towards the transverse plane

where p_x, p_y and p_z are the x, y and z components of the momentum. The projection of the momentum in the transverse plane is defined as transverse momentum, $p_{\mathrm{T}} = |\boldsymbol{p}| \sin\theta$ or

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \tag{2.2}$$

The measurement of rapidity (Equation 2.3) is preferred over θ because the difference in rapidity, Δy, is Lorentz invariant.

$$y = \frac{1}{2} \ln\left(\frac{E + p_z}{E - p_z}\right) \tag{2.3}$$

where E is the energy of the particle. Pseudorapidity is another important quantity defined as:

$$\eta = -\ln(\tan\theta/2) \tag{2.4}$$

For particles travelling close to the speed of light, E≫m (m is the mass of the particle), the pseudorapidity and the rapidity are equivalent. Since the precise measurement of the energy and momentum is more difficult than measuring the directions, pseudorapidity is considered instead of rapidity for highly relativistic particles. The distance between the particles, ΔR, in η-ϕ space is given by:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.5}$$

2.2.2. Inner Detector

The Inner Detector (ID) [82] is closest to the interaction point and is responsible for tracking of the charged particles emerging from the collision. The ID is immersed entirely in a 2 T magnetic field generated by a superconducting solenoid in order to measure precisely the momenta of the charged particles. It is composed of three subsystems - the pixel detector, the semiconductor tracker and the transition radiation tracker providing a coverage of $|\eta| < 2.5$ as shown in the r-z cross-sectional view (Figure 2.3). The particles interact with these layers and release a localised signal. These space points (*hits*) are recorded along the charged particle trajectory which after combination lead to reconstruction of *tracks* (Section 4.1). The momentum of the particle is determined by the curvature of the track. The track momentum resolution [83], $\sigma_{p_{\mathrm{T}}}$, of the ID as a function of the track transverse momentum p_{T} is given by

$$\frac{\sigma_{p_{\mathrm{T}}}}{p_{\mathrm{T}}} = 0.07\% p_{\mathrm{T}} \oplus 1\% \tag{2.6}$$

Table 2.1 summarizes the main characteristics of the ID subsystems.

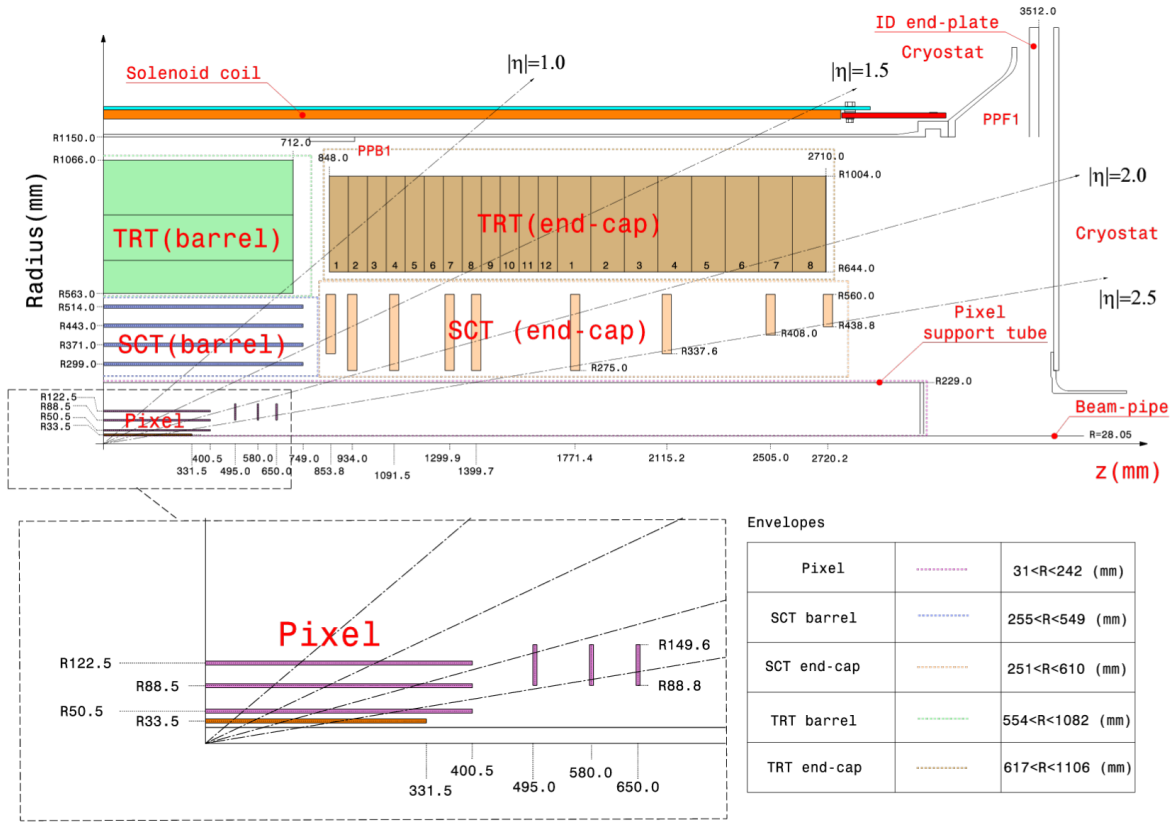


Figure 2.3.: The *r*-*z* cross-section view of a quadrant of the ATLAS inner detector for Run-2. The top panel shows the whole inner detector, whereas the bottom-left panel shows a magnified view of the pixel detector region. Compared to Run 1, the IBL (shown in red in the bottom-left panel) together with the new beam pipe are added in Run-2 [84].

Subsystems ID	Radius barrel layers (mm)	Element size	Intrinsic resolution
IBL	33.2	$50 \times 250\,\mu\text{m}^2$	$8 \times 40\,\mu\text{m}^2$
Pixel	50.5, 88.5, 122.5	$50 \times 400\,\mu\text{m}^2$	$10 \times 115\,\mu\text{m}^2$
SCT	299, 371, 443, 514	$80\,\mu$m	$17\,\mu$m
TRT	from 554 to 1082	$4\,\mu$m	$130\,\mu$m

Table 2.1.: Summary of the main characteristics of the ID subdetectors. The intrinsic resolution of the IBL and the Pixel is reported along *r*-ϕ and *z*, while for SCT and TRT is only along *r*-ϕ. For SCT and TRT the element sizes refers to the spacing of the readout strips and the diameter of the straw tubes, respectively [83].

Pixel Detector

The pixel detector is the innermost layer of the ATLAS tracking system which provides high-resolution measurements of parameters of charged particle tracks. It is made up of silicon pixel sensors which works on solid-state ionization principle. When a particle passes through the silicon layer, it creates electron-hole pairs which is proportional to the energy loss of the incident particle. The collected charges induce a signal in the pixel electrodes, allowing to determine the precise position of this particle.

During the long-shutdown of 2012-2015 (LS1) after the end of Run-1, a new barrel layer of silicon pixel detector, the Insertable B-layer (IBL) [85], was installed reducing the distance of the first sensitive layer to the interaction point from 5 cm to 3.3 cm. Hence, the pixel detector for Run-2 is comprised of four layers of barrel pixel detector and three pixel disks on each ends. The insertion of the IBL improved the tracking precision significantly due to a closer positioning from the interaction point.

The IBL sensors with a pixel size of $50 \times 250\,\mu\text{m}^2$ have an intrinsic measurement accuracy of $8\,\mu$m in the transverse plane (r-ϕ) and $40\,\mu$m along the z-axis. The outer three layers are composed of $50 \times 400\,\mu\text{m}^2$ sensors providing an intrinsic resolution of $10\,\mu$m in the r-ϕ plane and $115\,\mu$m along the z-axis [83].

As the pixel detector provides precise measurements of the track parameters, it contributes in distinguishing the b-hadron decays from other light-quark decays, essential for *b-tagging* discussed in Section 4.5.

Semiconductor Tracker

The second layer of the ID is the Semi-Conductor Tracker (SCT) which comprises of silicon micro-strip sensors arranged in four barrel layers and nine disks in each end-cap region. Each layer of SCT employs two sets of strips with a relative rotation of 40 mrad in order to obtain full three-dimensional position measurement. It is designed such that a charged particle crosses at least four layers of SCT modules. This provides an exemplary track reconstruction capability and better momentum resolution. Similar to the pixel detector, the SCT also works on the ionization principle for particle detection. It provides an intrinsic resolution of $17\,\mu$m in the azimuthal (r-ϕ) plane and $580\,\mu$m in the longitudinal direction [83].

Transition Radiation Tracker

The Transition Radiation Tracker (TRT) is the outermost component of the ID and is composed of drift tubes (straws) filled with a Xe-CO_2-O_2 (70-27-3%) gas mixture. Due to some gas leakage

before LS1, parts of the TRT gas system were filled with an Ar based mixture for Run-2. The straws are placed parallel to the beam axis in the barrel region and radially in the end-cap regions interleaved with polypropylene polymer acting as transition radiation material. The straws, 4 mm in diameter are made of a Kapton-based multilayer material functioning as cathode and contain 30 μm diameter gold-plated tungsten wires serving as anodes. When highly relativistic charged particles cross the boundary between these two mediums of different dielectric constants (CO_2 and polypropylene), X-ray photons are produced called transition radiation. The intensity of the radiation is proportional to the Lorentz factor, $\gamma = E/m$, of the incoming particle. For particles coming with the same momentum, γ is larger for lighter particles which implies that the electrons will radiate more photons than the pions. This feature aids in discrimination of ultra-relativistic electrons from charged pions.

The TRT provides an intrinsic resolution of 130 μm [83] in the azimuthal plane which is slightly worse than the silicon detectors but it provides more than 30 hits contributing to the precision tracking within $|\eta| < 2.0$.

Therefore, the TRT provides a combination of continuous tracking based on individual straws and also electron identification (Section 4.2) based on transition radiation from polypropylene polymer interleaved between the straws.

2.2.3. Calorimeters

The next detecting subsystems in the ATLAS detector are the calorimeters [86]. These are designed to measure the energy of the particles and aim to stop the particles by absorbing most of their energies. Depending on the particles to be measured, the calorimeter system is divided into two parts: Electromagnetic Calorimeter (ECal) for objects like electrons, photons etc. through their electromagnetic (EM) interactions and Hadronic Calorimeter (HCal) for hadronic particles through their strong interactions. In addition, Forward Calorimeters (FCal) are mounted on the end-caps to increase the longitudinal acceptance of the detector. The energy resolution of the calorimeter is summarised in Table 2.2 as determined from the test beam results.

Electromagnetic Calorimeter

The ECal uses Lead-Liquid Argon (Pb-LAr) sampling calorimeter technique with accordion-shaped electrodes. LAr is the active layer, Pb acts as an absorber and the accordion geometry of the readout system provides full range in ϕ without any crack. The ECal is organised in two

Calorimeter components	Resolution
Electromagnetic	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
Hadronic	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$
Forward	$\sigma_E/E = 28.5\%/\sqrt{E} \oplus 3.5\%$ (FCal1) $\sigma_E/E = 94\%/\sqrt{E} \oplus 7.5\%$ (FCal2 and FCal3)

Table 2.2.: The calorimeter energy resolution of the ATLAS detector as discussed in [87], [88] and [89].

identical half-barrels separated by few mm at $z = 0$ covering up to $|\eta| < 1.475$ and two end-caps covering $1.375 < |\eta| < 3.2$. An additional thin LAr presampler covering $|\eta| < 1.8$ allows corrections for energy losses for particles that start showering before entering the calorimeter. The transition between the barrel components and end-caps leads to poorer performance of EM object reconstruction in the region $1.37 < |\eta| < 1.52$.

The ECal barrel modules are constituted by three layers with varying depth and different η, ϕ granularity as shown in Figure 2.4a. The first layer has very fine fragmentation in η in order to provide discrimination between single photon and two overlapping showers coming from π^0 decay by resolving their impact points. The second layer accounts for most of the thickness of the ECal and is responsible for absorbing most of the EM showers. The last thin layer with even coarser granularity is designed to collect the tail of the EM showers and to separate it from hadronic ones. The design energy resolution of the ECal is $\sigma_E/E = 10\%/\sqrt{E} \bigoplus 0.7\%$ [87].

Hadronic Calorimeter

The HCal uses tile sampling calorimeter technology for the central and extended barrel region covering $|\eta| < 1.7$ and LAr technology for the end-cap region ($1.5 < |\eta| < 3.2$). The tile calorimeter is composed of a steel absorber combined with scintillating tiles as active material. The signal from the scintillating tiles is read out by wavelength shifting fibres and directed to photomultipliers that convert the light into current pulses as depicted in Figure 2.4b. The end-caps consist of copper plates as absorbing material interleaved with LAr gaps as active material for its robustness against high radiation which is expected in forward region. The hadronic calorimeter is designed to give an energy resolution of $\sigma_E/E = 50\%/\sqrt{E} \bigoplus 3\%$ [88].

Forward Calorimeter

The calorimeter is completed by the FCal system [89] which provides EM and hadronic energy measurements in $3.0 < |\eta| < 4.9$ range. First layer of forward detector (FCal1) is the electromagnetic component which also uses LAr as the active medium but copper is chosen as the absorbing medium to facilitate heat removal. The second and third layer of the FCal

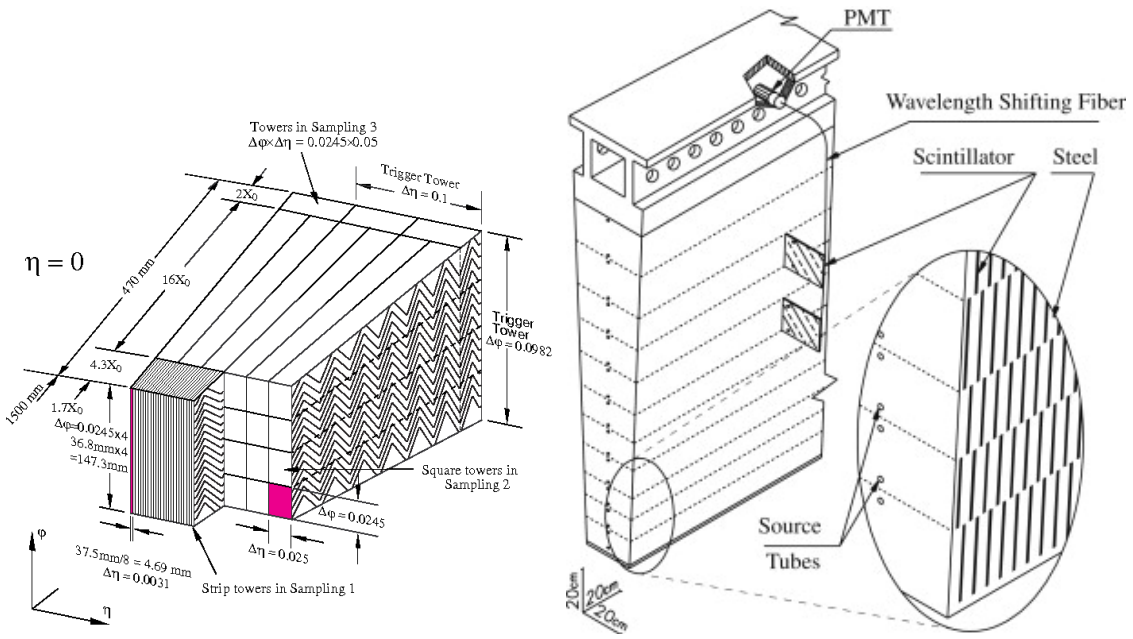


Figure 2.4.: (a) Sketch of the three-layer segmentation of the ECal with their respective η, ϕ granularities and (b) schematic of a barrel segment of the HCal [78].

(FCal2 and FCal3) optimized for hadronic measurements use tungsten as absorbing material to accommodate the hadronic showers in a limited volume. As quoted in Table 2.2 the resolutions of the FCal detectors are not as good as other parts of the calorimeter because of the coarser segmentation. A major objective is to contribute to the hermetic geometry of the ATLAS detector which is crucial for the measurement of missing transverse momentum.

2.2.4. Muon Spectrometer

The Muon Spectrometer (MS) [90] forms the outermost layer of the ATLAS detector to detect muons which escape all inner layers of the detector. The fabrication includes three layers of barrel ($|\eta| < 1.0$) and four layers of end-caps extending up to $1.4 < |\eta| < 2.7$, leaving a transition region between $1.0 < |\eta| < 1.4$. The muon system is immersed into a magnetic field of 1 T, generated by air-core toroidal magnets for each layer. The air-core magnet concept minimises the density of material traversed by the muons and hence reduces the chances of multiple scattering. This improves the tracking resolution.

The MS consists of four components: the Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC) as two precision muon trackers and two triggering sub-systems namely, Resistive Plate Chambers (RPC) and Thin Gap Chambers (TGC). The majority of the precision

tracking chambers are constituted by MDT that cover up to $|\eta| < 2$ which provide a precision measurement of the track coordinates and muon momenta in the r-z plane. The active material is an aluminium tube filled with an Ar-based gas mixture with a tungsten wire at the centre acting as anode. Muons are detected with a single hit with a resolution of 80 μm. The accuracy is improved by employing three to eight layers of drift tube in each chamber. CSCs are used for the precise measurement of momenta in the region $2.0 < |\eta| < 2.7$ designed to cope up with the larger number of traversing particles in this region. The region $|\eta| < 1.05$ is equipped with RPCs while $1.05 < |\eta| < 2.4$ contain TGC trigger systems. These systems provide trigger information with faster time of response and also supplement the tracking chambers by measuring the coordinates in the η-ϕ plane. The designed p_{T} resolution of the MS is $\sigma_{p_{\mathrm{T}}}/p_{\mathrm{T}} = 0.29/p_{\mathrm{T}} \oplus 0.043$ [91].

2.2.5. Trigger System

With a bunch spacing of 25 ns in the LHC Run-2, the pp interaction rate is 40 MHz. Storing all 40 million events per second would require data recording at a rate of roughly 60 TB/s. A continuous extraction of collision data without any filter is not feasible due to the current limitation in data transfer speed and data storage. In addition, only a fraction of these events are interesting for physics analyses. Hence, the ATLAS Trigger and Data Acquisition System (TDAQ) is very crucial to manage the data flow and select the interesting events. This system underwent several modifications during LS1 to deal with the increased instantaneous luminosity, decreased bunch spacing from 50 ns to 25 ns and increased pileup. For Run-2, it consists of a hardware-based first level trigger (L1) and a software-based high level trigger (HLT).

The L1 trigger uses a coarser information from the calorimeter system and trigger chambers of the MS and reduces the event rate from 40 MHz to 100 kHz, with a processing time of 2.5 μs. It determines Regions of Interest (RoI) defined as the η-ϕ region of the detector where the system identifies high p_{T} jets, leptons or photons. The RoI is sent to the HLT that performs a partial *online* event reconstruction through an analysis of all the available data with fine granularity. It reduces the rate to 1 kHz with an average decision time of 200 ms. If an event is accepted by HLT, the data is stored to perform full *offline* event reconstruction.

2.3. Luminosity Measurement and Data Sample

The instantaneous luminosity measures the ability of an accelerator to produce a required number of interactions. It is the proportionality factor between the number of events produced per second (dN/dt) and the cross-section (σ) for the physics process (in case of LHC it is the

pp inelastic cross-section).

$$\frac{dN}{dt} = \mathcal{L}\sigma \tag{2.7}$$

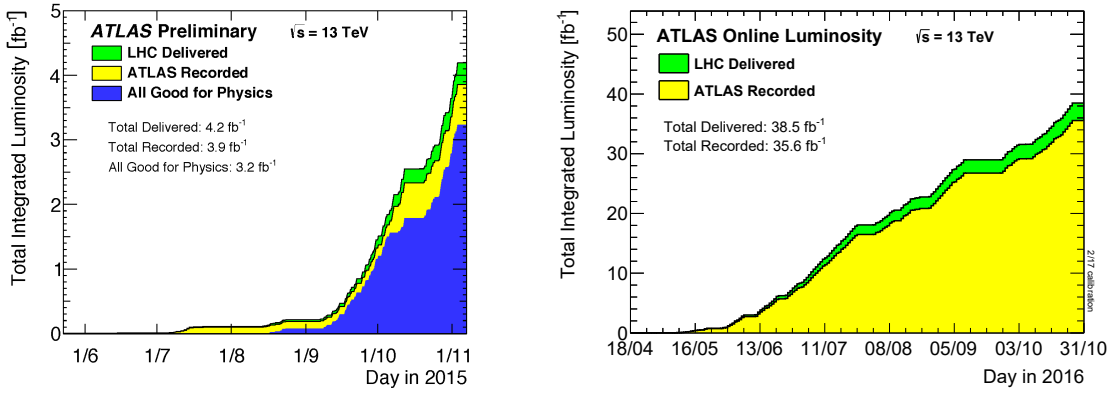
The unit of the instantaneous luminosity is therefore $\mathrm{cm}^{-2}\mathrm{s}^{-1}$. It can be expressed as:

$$\mathcal{L} = f n_b \frac{N_1 N_2}{2\pi \Sigma_x \Sigma_y} \tag{2.8}$$

where f is the beam revolution frequency, n_b denotes the number of bunches colliding per revolution, N_1 and N_2 are the number of protons in the two bunches. Σ_x and Σ_y are the beam spread in the x and y directions, respectively. The integrated luminosity (L) is the integral of the delivered instantaneous luminosity over time.

$$L = \int \mathcal{L} dt \tag{2.9}$$

ATLAS measures the delivered luminosity with two dedicated detectors: Beam Conditions Monitor (BCM) [92] and LUminosity Cherenkov Integrating Detector-2 (LUCID-2) [93]. LUCID-2 is the primary source of online monitoring of instantaneous luminosity by detecting the inelastic pp collisions using photomultiplier tubes and plays a crucial role in the offline determination of the integrated luminosity. It is closest to the interaction point located at a distance of $z = \pm 17\,\mathrm{cm}$. The BCM consists of diamond sensors located symmetrically around the beam axis at $z = \pm 184\,\mathrm{cm}$ and $r = 55\,\mathrm{mm}$. In addition to the bunch-by-bunch luminosity measurements, the BCM also monitors the stability of LHC beam.



(a) **(b)**

Figure 2.5.: The total integrated luminosity delivered by the LHC (green) and recorded by the ATLAS detector (yellow) in the year 2015 (Figure 2.5a) and 2016 (Figure 2.5b). For 2015, the figure also includes the data certified to be good for physics analyses (blue) [94].

This analysis uses the dataset collected by the ATLAS detector during the time period of 2015-2016. The LHC operates at $\sqrt{s}$ = 13 TeV in Run-2 (scheduled for 2015-2018) reaching the

peak instantaneous luminosity of $5 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ in 2015 and $13.8 \times 10^{33}\text{cm}^{-2}\text{s}^{-1}$ in 2016 [94]. Figure 2.5 presents the total integrated luminosity delivered by the LHC and recorded by the ATLAS detector in 2015+2016 which corresponds to $39.5\,\text{fb}^{-1}$ in total, whereas $36.1\,\text{fb}^{-1}$ of data is used for the physics analyses. This is due to the requirement that the detector must satisfy standard quality criteria which means all the detector components must be fully operational during the data taking. This list is called the Good Run List (GRL). Events not listed in the GRL are not taken for the analysis.

The preliminary uncertainty related to the combined 2015+2016 data corresponding to a total integrated luminosity of $36.1\,\text{fb}^{-1}$ is 2.1% which is considered for the measurements in this analysis. It is derived from a methodology detailed in Reference [95].

Along with the interesting hard interaction, additional pp interactions also occur which are referred to as *pileup*. The additional interaction can occur in the same bunch crossing due to the high instantaneous luminosity called in-time pileup or with the surrounding bunches due to small bunch-spacing which is called out-of-time pileup. The level of pileup is computed by considering the average number of interactions per bunch crossing, $\langle\mu\rangle$. Figure 2.6 shows the distribution of $\langle\mu\rangle$ for 2015 and 2016 ATLAS pp datasets. It can be seen that increased luminosity in 2016 results in higher number of interactions per bunch crossing than in 2015.

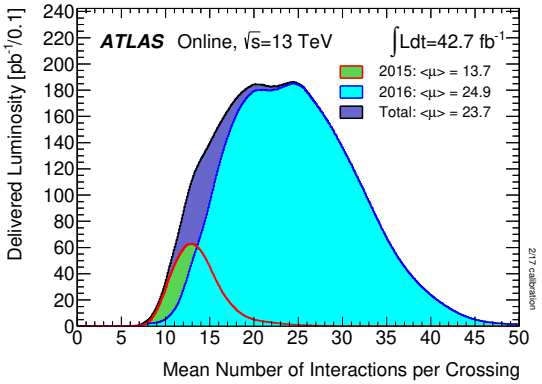


Figure 2.6.: Distribution of average number of pp interactions per bunch crossing for the year 2015 and 2016 for the ATLAS detector [94].

It is a difficult task to produce the same pileup profile in MC simulation as obtained from data. Therefore, a reweighting is performed on the MC pileup distribution that reweights the MC to reproduce the data pileup profile. The uncertainty in the pileup reweighting is estimated by comparing the distribution of the number of primary vertices in the MC simulation to the data as a function of instantaneous luminosity. The differences between these distributions are adjusted by scaling $\langle\mu\rangle$ in the MC and an uncertainty of $\pm 1\sigma$ is assigned to these scaling factors. The pileup weights are recalculated after varying the scale factors within their uncertainties.

Chapter 3.

Monte Carlo Simulation

The physics processes involved in an analysis, like the one presented here, are modelled using one of the most popular computational algorithms called the Monte Carlo (MC) generators. MC is a stochastic technique utilizing repeated random sampling of variables and provides the result in the same format as the measured ones. In addition it allows full access to the evolution of each event from pp collision to the reconstructed objects. It simulates both the hard scattering processes predicted using perturbation theory and the soft interactions mainly characterised by non-perturbative effects, described by phenomenological models. There are mainly two primary steps in the simulation: event generation which includes matrix element generation followed by parton shower and hadronization modelling and the simulation of detector response as shown in Figure 3.1.

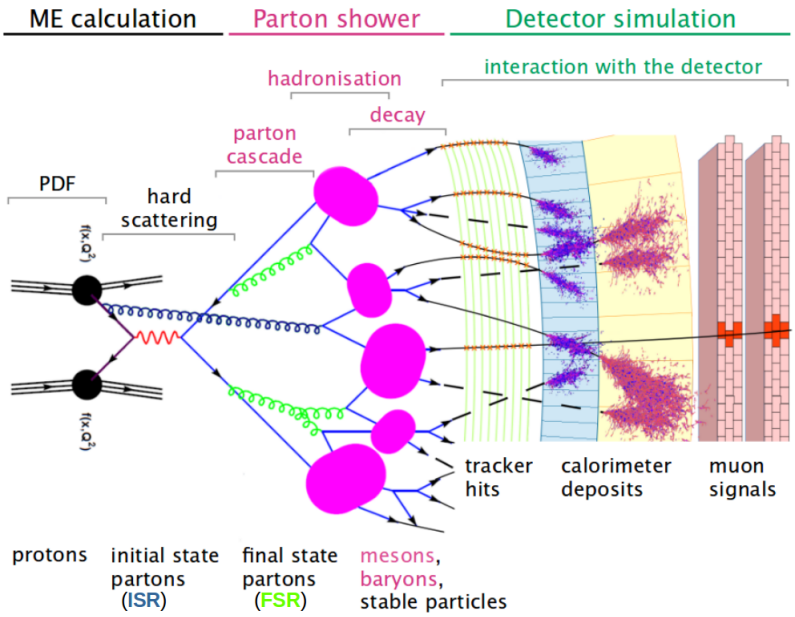


Figure 3.1.: Schematic of a typical pp collision event with different steps involved in event generation chain taken from [96] and edited.

This chapter describes the important steps of the full event simulation chain. Section 3.1 gives an overview of different types of generators used in the event generation process. As $t\bar{t} + b$-jets is rather a complex process, Section 3.2 explains the approaches used by different generators in order to handle the production of additional b-jets. Section 3.3 mentions the final step of the simulation procedure i.e. the simulation of the detector response. Section 3.4 describes the generators used to simulate the signal and background processes relevant in this analysis. The choice of a particular MC generator to describe a process leads to a bias because of the generator specifications like the scale choice which leads to modelling uncertainties. Section 3.5 discusses the modelling uncertainties which are included in the results of this thesis.

3.1. Event Generation

Monte Carlo simulated events are used to compare the theoretical expectations to the experimental data from pp collisions inside the ATLAS detector. It simulates all the events completely, starting from the high-energy collision down to the level of individual stable particles. They are broadly classified as the matrix element (ME) generators and General Purpose Monte Carlo (GPMC) generators. These two generators are then combined using different matching and merging algorithms to simulate precisely the hard and soft regimes of high-energy pp collision.

3.1.1. Matrix Element Generators

The ME generators compute the hard scattering process up to a certain order in perturbation theory and integrated over the PDFs to address all possible initial partonic states. As discussed in the theory chapter, adding higher orders increases the precision of the calculation. Figure 3.2a represents the LO diagram for the production of $t\bar{t}$ process and Figure 3.2b shows a one-loop NLO correction to it. Figure 3.2c shows a real emission of a gluon which contributes to higher order multi-leg calculation with LO precision. Addition of more legs is good for the kinematic distributions and is important to deal with the complexity of a process such as production of additional b-jets in this analysis. Since the higher order terms are defined with respect to the LO diagram, Figure 3.2d corresponds to NLO correction if the process of interest is $t\bar{t}$ production with an additional gluon i.e. Figure 3.2c. It should be noted that Figure 3.2d with one-loop and one real emission can also be considered as NNLO correction to Figure 3.2a. Usually the number of additionally produced partons are limited in the hard scattering calculation due to computational speed and complexity.

The ME generators are only capable of generating the hard scattering processes and everything

else remaining in the event is done by one of the GPMC generators feeding the ME. Typical ME generators used for the hard process are:

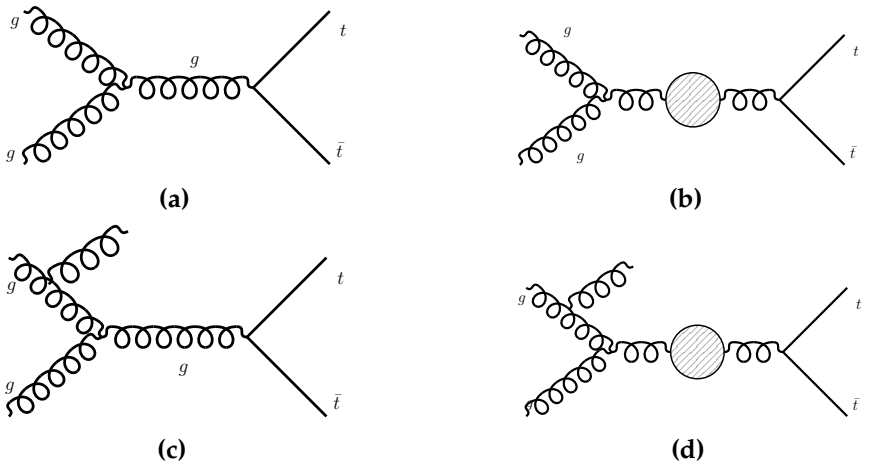


Figure 3.2.: Examples of Feynman diagrams for (a) $t\bar{t}$ process at LO and (b) NLO loop-correction to it. (c) shows a real emission in $t\bar{t}$ with (d) as NLO loop-correction correction to it.

- **POWHEG** [97] stands for the Positive Weight Hardest Emission Generator, is an event generator computing ME at NLO parton-level with both real and virtual corrections in pQCD. This formalism builds each event by producing $2 \to 2$ or $2 \to 3$ partons in hard scattering matrix element. For the process relevant in this analysis, the $t\bar{t}$ event is simulated at NLO accuracy. The generated events need to be interfaced with the parton shower (PS) like PYTHIAand HERWIG using the POWHEG method [97].

- **MG5_aMC@NLO** [98] provides an automated computation of the tree-level and NLO matrix element. Unlike the POWHEG method, in MC@NLO, it is possible to assign negative weights to a small fraction of events. The negative weights account for the cancellation of interfering diagrams. PYTHIAis used to interface the PS in this thesis.

- **POWHEL** [99] is an event generator which is a variation of POWHEG. The difference lies in the technicality for the computation of the amplitudes of matrix elements. POWHEL (POWHEG + HELAC) uses HELAC-NLO codes [100–102] to compute the amplitude required as input by POWHEG-BOX. PYTHIAis interfaced for the PS part.

3.1.2. General Purpose MC Generators

The GPMC generators are capable of simulating the entire pp collision from the hard scattering ME followed by the parton shower and hadronization plus the underlying events (UE). However, except SHERPA, other GPMC generators are used only to feed the ME generators. Therefore, after the ME calculation shown in black in Figure 3.1, the PS models recursively split

the radiated gluons reducing the energy scale of the process. The PS starts at some high energy scale determined by the hard scattering and continues until a scale of order 1 GeV, at which point a hadronization model takes over to convert the resulting partons into hadrons. Common hadronization models are the cluster model [103] and the Lund string model [104, 105]. The underlying events from proton remnants and multi-parton interactions (MPI) as discussed in Chapter 1 and the pileup (Section 2.3) are also modelled in GPMC generators. Some widely used examples are:

- **PYTHIA** [106] can provide hard scattering simulation for various processes limited to $2 \to 1$ or $2 \to 2$ partons at LO, or can be interfaced with the output of external ME generators to generate the PS, hadronization and UE. In this analysis, PYTHIAfulfils the later purpose only because the ME generators can provide NLO accuracy in the calculation. The parton shower emissions accounting for the ISR/FSR are ordered in transverse momentum. The hadronization modelling in PYTHIArelies on the Lund string model. The underlying events also include the MPI events, are modelled using different tunes [107, 108] for version 6 and version 8 of PYTHIA.

- **HERWIG** [109, 110] is another GPMC generator providing an interface to ME generators in this analysis. Here the parton shower emissions are ordered in opening angle. It implies that for a parton a branching into products x and y, the scale is given by $Q^2 = 2E_a^2(1 - \cos\theta_{xy})$ where the energy of parton a equals E_a and θ_{xy} is the angle between x and y. The angle between the products of successive branching decreases as the shower evolves. Hadronization is based on the cluster model while the UE is modelled using minimum-bias interactions.

- **SHERPA** [52] is capable of providing NLO hard scattering calculations. Unlike the previous two GPMC, SHERPA is used for the simulation of both hard and soft interactions. The Simulation of High-Energy Reaction of PArticles (SHERPA) is a steering module that handles all the steps of simulation by various programs initialized and controlled by the SHERPA community itself. The PS is matched to the ME based on the Catani-Seymour dipole formalism [111, 112]. It provides a better description of high jet multiplicity compared to sother generators.

3.1.3. Tuning

The QCD processes are very well known where the perturbation theory applies. Low energy effects or soft processes such as hadronization and UE use phenomenological models to describe physics processes. The MC event generators consist of a number of free parameters mostly related to modelling of these soft and non-perturbative physics which cannot be

constrained by first principle and need to be determined from data. This process is referred to as tuning. The free parameters can regulate the ISR/FSR contributions, multi-particle interactions, beam remnants, pileup and so on to provide a good description of data. The tuning is applied to each subset of models separately using the clean LEP [113] measurements for the hadronization and FSR parameters to Tevatron and LHC data to constrain the modelling of the ISR and MPI. PYTHIAutilizes the A14 [107] while Herwig uses the H7UE [110] set of tunable parameters.

3.2. Calculation of $t\bar{t}b\bar{b}$ in MC Generators

The presence of several partons in the final state of the $t\bar{t}$ + b-jets process, especially the additional b-jets, leads to complications in event generation. Therefore, it is of great relevance to understand how different generators handle these additional b-jets. Consider a process

$$p_1 + p_2 \longrightarrow s + n \text{ - jets} \tag{3.1}$$

where p_1 and p_2 are two protons, s is a set of final-state stable particles and n is the number of additional jets to be studied in the final-state (not to mention that the number of jets in hadron colliders are huge). The process can be computed in several ways:

- In order to do a measurement one straightforward approach is that of using ME with $n = 0$ jets and letting the shower generate all the emissions. In other words, only the process $pp \rightarrow t\bar{t} \rightarrow WWb\bar{b}$ is computed from the matrix element at NLO and the additional b-quarks are simulated using parton shower generators. A drawback in this procedure rises because the PS simulates only the softer emissions well not the hard jets. Therefore the additional b-quark kinematics are calculated correctly in soft limits only. However, these emissions are mostly expected to be produced in this region, so this approximation is also considered. The MC generators used in this analysis which lie under this approximation are POWHEG + PYTHIA8, POWHEG + HERWIG7 and MG5_aMC@NLO + PYTHIA8.

- A few generators provide another approach where some additional partons can also be calculated in the ME. For the case where $m \leq n$, these generators allow the calculation of $t\bar{t}$ + m-jets from the ME and let the remaining come from the PS. This method increases the accuracy of calculation. One variation of the SHERPA generator, specifically referred to as Sherpa 2.2 $t\bar{t}$ in this analysis, utilizes this method. Here, only additional gluon radiation is calculated in ME but not the splitting to a $b\bar{b}$ pair. Hence, these simulations do not provide the full LO calculation of the $t\bar{t}b\bar{b}$ process.

- Some predictions are also available which calculate the full $pp \rightarrow t\bar{t}b\bar{b}$ process in the matrix element at NLO. Generators like Sherpa 2.2 $t\bar{t}b\bar{b}$, POWHEL + PYTHIA8 $t\bar{t}b\bar{b}$ and POWHEG + PYTHIA8 $t\bar{t}b\bar{b}$ incorporate this feature and are expected to provide most accurate predictions.

3.2.1. Matching and Merging

It is seen that the matrix element generation relies on the pQCD calculation at fixed order and the parton shower includes contributions from logarithmic terms to a few orders in the collinear limit. Since the two approaches are complementary, a proper and consistent combination of these theories are required to completely describe an event in pp collision.

The generators like POWHEG and MG5_aMC@NLO which contain $t\bar{t}$ at NLO in ME have three contributions: one is the LO diagram (Figure 3.2a), one internal loop NLO correction (Figure 3.2b) or one real emission NLO correction (Figure 3.2c). Now, if the PS is interfaced with ME to the LO contribution, one obtains the same configuration as Figure 3.2c where the radiation is coming from PS. This leads to double counting of the same configuration as shown in Figure 3.3 (b) and (c). There are two methods for matching NLO calculations with the PS namely MC@NLO [114] and POWHEG [97]. These methods take the highest p_{T} emission from the NLO ME with the shower approximation subtracted and all other emissions from the PS making them reliable mostly in collinear limits only.
For the generators like Sherpa 2.2 $t\bar{t}$ with higher particle multiplicities in final state i.e. having $t\bar{t}$ + 0,1-jet in ME at NLO, the situation is even more complicated. The configurations in Figure 3.3 (a) and (b) are already part of ME calculation for $t\bar{t}$ +0-jet and in addition Figure 3.3 (d) is a LO contribution to $t\bar{t}$ +1-jet. All three diagrams have the same configuration which causes merging of LO+PS calculations to be very difficult for different final state particle multiplicities. The basic idea used in this calculation is to divide the phase space into two regions with the use of a transition scale. Say the transition scale in terms of energy is q_t^2, the energy of hard-scattering process is Q_0^2 and the low energy where hadronization takes over is q_0^2. The PS evolution can occur at any energy scale q^2 varying from Q_0^2 all the way to q_0^2. In the merging of LO+PS scheme, for $q^2 > q_t^2$, the observables are taken from matrix element while for $q^2 < q_t^2$ the contributions are taken from the PS. The evolution of PS occurs iteratively using Sudakov form factor and stops for the value of $q^2 \leq q_0^2$. The scale can be any parameter like energy or transverse momentum. The schemes which utilize this approach are CKKW [115], CKKW-L [116] and MLM [117].

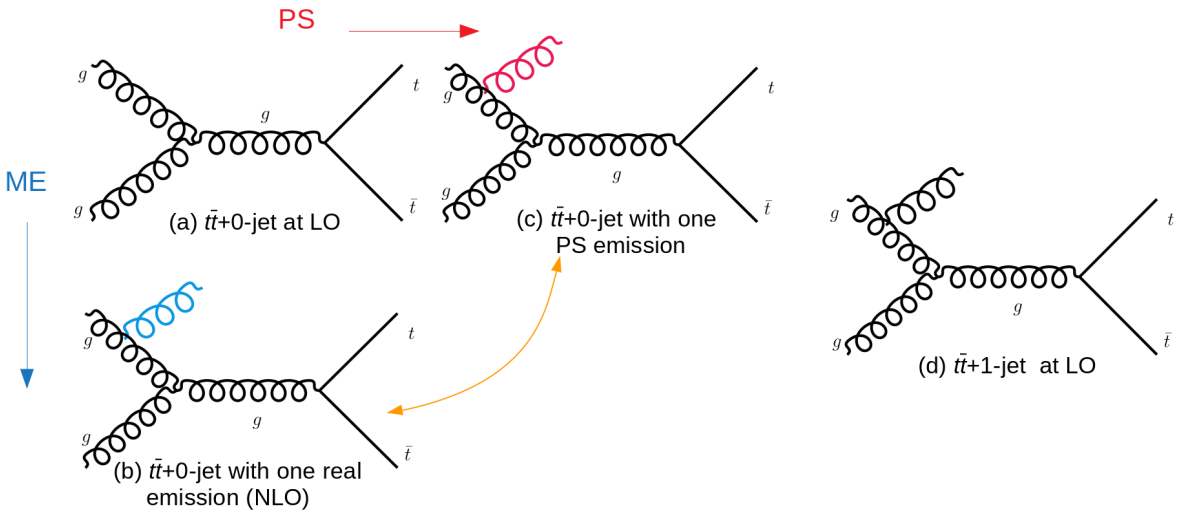


Figure 3.3.: A schematic representation of double counting issue in matching and merging. Figure (a) shows a LO $t\bar{t}$ +0-jet Feynman diagram and (b) shows the real emission NLO correction to (a). The diagram (c) illustrates a possible configuration when the event in Figure (a) is propagated through a parton shower. The same configuration of (b) and (c) demonstrates the double counting problem. In addition, Figure (d) is a LO contribution for $t\bar{t}$ +1-jet process. Therefore, merging LO+PS with different particle multiplicities in final state also results in double counting of certain configurations. For simplification, no NLO loop corrections are shown above.

3.2.2. 4 and 5-Flavour Schemes

The physics process covered in this analysis involves the b-quarks in the initial and final state of the calculation. Due to the non-negligible mass of the b-quark these calculations can be performed in two different schemes. The first scheme is the 4-flavour scheme (4FS) or "massive" scheme while the other one is 5-flavour scheme (5FS). Both approaches come with some pros and cons depending upon the observable under study. The flavour schemes are introduced below based on reference [75].

Four-Flavour Scheme

In the 4FS, it is assumed that since the b-quark is significantly heavier than the proton mass it can only appear in the final state. In this case the PDF consists of only four flavours (u, d, c and s) and the PDF for b-quark is set to zero. Consequently, the b-quark does not enter in the evolution of the PDFs or in the computation of the running coupling constant. The 4FS is more reliable for the observables involving the b-quark kinematics as it considers the mass of the b-quark explicitly. However, the occurrence of large logarithms $\left(\ln \frac{Q^2}{m_b^2}\right)$ which are not resummed may lead to an underestimation of the total inclusive cross-section and displays

higher sensitivity to the scale choice as compared to the 5FS [75]. Moreover, the presence of massive b-quark and multiple external legs make the computation complicated.

Five-Flavour Scheme

The 5FS also referred to as "massless" scheme is a natural choice given the high-energy of the incoming protons which is huge as compared to the mass of b-quark. In this case the PDFs contain 5 flavours (u, d, c, s and also b), so the b-quark enters in the PDF evolution as well as in the coupling constant. The bottom quark can appear both in initial and final state. The logarithms are resummed by the PDF evolution leading to a more accurate prediction of the total inclusive cross-section and improved stability of the calculation with varying scales. Also, the reduction in the number of scales involved and the number of external legs facilitates higher order calculation as compared to 4FS. The shortcoming of this scheme is that it gives less precise description of the b-quark kinematics. It can also not describe the phase space region where both the b-quarks are soft or merged in a single jet.

3.3. Detector Simulation

The event generators model the physics of different processes and interactions occurring among the particles in pp collision. The output from the event generators is a list of four-vectors of all the stable particles produced in the event. In order to compare the output from the MC generator with the experimental data recorded by the ATLAS detector, the generated events are propagated through an ATLAS like infrastructure using the software package GEANT4 (GEometry ANd Tracking) [118, 119]. Accurate simulation of the interaction of particles with the detector also allows determination of the detector response and efficiencies. For the accuracy of the simulation, information from two broad perspectives of detector is required: one is the description of the detector volume in terms of the position, geometry, dimension and material composition, while the second is the information about the detector real-time response i.e. condition of the detector in each run like dead channels or misalignments. The output after the detector simulation has the same format as the real data.

It is established that the process of event generation undergoes several steps which gives the opportunity to study the process at various levels. These levels can be classified as:

- **Parton Level:** The matrix element calculation of the hard interaction gives the partons. Therefore, to study the partons, the stage right after the hard scattering is referred to as *parton-level*. Referring back to Figure 3.1, the part displayed in black colour under the ME

calculation is the parton-level. At this level the particles do not suffer from any radiation or showering. Particle behaviour at this level is strongly dependent on the choice of the ME generator.

- **Particle Level:** The final output of the event generation process which give the four-vectors corresponding to the stable particles is called the *particle-level*. The resulting particles after all the radiations, hadronization and decays with a life-time of $\sim 3 \times 10^{-11}$ s are stable particles. This stage is commonly used to perform the study due to the fact that they are less dependent on the choice of MC generators. This level is frequently referred to as the *truth level* in this analysis.

- **Detector Level:** Fully reconstructed objects which can directly be compared with the data is called *detector-level* or *reconstruction-level*. It is obtained after the full simulation process, that is, the stage after the events are reconstructed in the detector. It is easily perceivable that the data is obtained only at the detector-level and to perform the study at particle-level, an unfolding procedure is required which is discussed in this thesis.

3.4. Simulated Samples

Monte Carlo event generators are used to simulate the signal process and background composition in the selected data sample. The signal and background processes are described in detail in Section 5.1. In this analysis the MC simulations are also used to determine correction factors for the efficiency and acceptance effects of the detector and to estimate systematic uncertainties. For the modelling of the signal process, several generators are used, having different perturbative accuracy and covering a range of choices for the parton shower, hadronization and other generator effects (Section 3.4.1). A set of simulated theoretical predictions are also used to compare with the resulting unfolded data. These state-of-the-art MC samples are also explained below with other setups used to generate the samples and are also summarized in Table **??**.

3.4.1. Signal Modelling

Signal refers to the process of interest which includes $t\bar{t}$, $t\bar{t}H$ and $t\bar{t}V$ processes in this analysis.

Nominal $t\bar{t}$ Signal Sample

The nominal signal sample is generated using the POWHEG-BOX generator, version 2 [97, 114, 120] which offers the NLO matrix element for the $pp \to t\bar{t}$ process [121]. The NLO

NNPDF3.0NLO PDF set is used for the matrix element calculation. Additional b-quarks are modeled from the parton shower. The parton shower, fragmentation, and the underlying event are simulated using PYTHIA8.210 [106]. The PDF set NNPDF2.3LO [122, 123] and the corresponding A14 set of tunable parameters [107] are used for the parton shower. The renormalisation and factorisation scales are set to:

$$\mu = \sqrt{m_t^2 + p_{\mathrm{T},t}^2}\,. \tag{3.2}$$

where m_t and $p_{\mathrm{T},t}$ are the mass and transverse momentum of the top quark. The h_{damp} parameter, which controls the p_{T} of the first additional emission beyond the Born configuration, is set to 1.5 times the top mass. The cross-section is normalised to NNLO+NNLL order.

$t\bar{t}V$ and $t\bar{t}H$ Samples

Processes involving the production of a W, Z or Higgs boson in addition to a $t\bar{t}$ pair also contribute to the signal. Since these bosons can decay to b-quarks, they provide the same final state as $t\bar{t}$ plus additional b-jets. These processes are simulated using the MG5_aMC@NLO generator [98, 124] at NLO accuracy in the ME. PYTHIA8 with the A14 tune is used to simulate the parton shower, fragmentation, and the underlying event. Overlap between the matrix element and additional radiations in the parton shower are removed using the CKKW-L merging scheme. The NNPDF3.0NLO PDF set is used in the matrix element calculation while the NNPDF2.3LO PDF set is used in the parton shower.

3.4.2. Alternative $t\bar{t}$ Samples

Signal modelling uncertainties originate due to a particular choice of matrix element, the parton shower and hadronization model, and the choice of the PDF set used in the simulation. In addition, the modelling of the initial- and final-state QCD radiation (ISR/FSR) also accounts for ratiation systematic uncertainties. Alternative $t\bar{t}$ samples are generated to investigate and assess these uncertainties and to compare with unfolded data.

Radiation Systematic Samples

In order to estimate the effect of QCD radiation, two samples using POWHEG+PYTHIA 8 namely POWHEG+PYTHIA 8 (RadLo) and POWHEG+PYTHIA 8 (RadHi) are generated. These samples are produced by varying the factorisation and renormalisation scales by a factor of 2 (0.5) for less (more) parton-shower raditaion [125] using low (high) radiation variations of the A14 tune and an h_{damp} value of $1.5m_t$ ($3.0m_t$).

Sample to Evaluate Shower Systematics

In order to evaluate effect of the choice of parton shower algorithm used in the analysis, POWHEG samples were produced with the parton shower changed to HERWIG7 [109, 110] (v7.01) using the H7UE set of tunable parameters.

Sample to Evaluate Generator Systematics

In order to estimate the effect of the matrix element calculation, samples were produced with a different generator, namely SHERPA 2.2.1 [52]. It models zero and one additional parton emission at NLO and up to four additonal partons at LO accuracy, using the MEPS@NLO prescription [126]. The NNPDF3.0NNLO PDF set is used and 5FS treatment is applied. The calculation uses its own parton shower tune i.e. developed by the SHERPA community.

Additional Sample for Comparison and to Calculate the $t\bar{t}$ +V/H Fraction

One more $t\bar{t}$ sample is generated using MG5_aMC@NLO [98] (v2.3.3), interfaced to PYTHIA 8.210 to compare with the unfolded results. It also uses the NNPDF3.0NLO PDF set with the matrix element calculation and the NNPDF2.3LO PDF set in the parton shower. The A14 tune is used for PYTHIA.The fraction of $t\bar{t}$ +V / H events in the $t\bar{t}$ events is also computed using this sample.

3.4.3. Predictions from $t\bar{t}b\bar{b}$ Samples

This section describes the MC samples in which the complete $t\bar{t}b\bar{b}$ process is calculated in the matrix element whereas the previous Section 3.4.2 demonstrated the samples in which the additional $b\bar{b}$ pair is described by the parton shower. The renormalisation scale, μ_{R}, is set to

$$\mu_{\mathrm{R}} = \prod_{i=t,\bar{t},b,\bar{b}} E_{\mathrm{T}_i}^{\frac{1}{4}}$$

where E_{T_i} is the transverse energy of the parton i in the partonic final state. The factorisation scale, μ_{F}, is set to

$$\mu_{\mathrm{F}} = H_{\mathrm{T}}/2 = \frac{1}{2} \sum_{i=t,\bar{t},b,\bar{b},j} E_{\mathrm{T}_i}$$

where j refers to the additional QCD-radiated partons at NLO. Following four state-of-the-art MC samples with $t\bar{t}b\bar{b}$ ME are used only for the comparison with unfolded data.

SHERPA (4FS)

A dedicated sample of $t\bar{t}b\bar{b}$ events is generated using SHERPA+OPENLOOPS [19]. $t\bar{t}b\bar{b}$ matrix elements are calculated with massive b-quarks at NLO, using the COMIX [127] and OPENLOOPS [128] matrix element generators, and merged with the SHERPA parton shower, tuned by the authors [112]. The four-flavour NNLO NNPDF3.0 PDF set is used. The resummation scale, μ_{Q}, is set to the same value as μ_{F}.

The following three predictions are based on the POWHEG method, and use the PYTHIA8 parton shower with the same parton shower tune and the same matching settings as the nominal POWHEG+PYTHIA 8 sample, with the exception of the h_{damp} parameter, which is set to the same value as the factorisation scale, i.e. $H_{\mathrm{T}}/2$.

POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)

This sample is generated using the POWHEL generator [20], where the matrix elements are calculated at NLO assuming massless b-quarks (5FS) and using the five-flavour NLO NNPDF3.0 PDF set. Events are matched to the PYTHIA8 parton shower using the Powheg method.

POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)

To generate this sample, the POWHEL generator is used where the matrix elements are calculated at NLO with massive b-quarks (4FS). The four-flavour NLO NNPDF3.0 PDF set [21] is used. Events are matched to the PYTHIA8 parton shower using the Powheg method.

POWHEG+PYTHIA 8 (4FS)

A dedicated sample of $t\bar{t}b\bar{b}$ events using the POWHEG generator is also used where $t\bar{t}b\bar{b}$ matrix elements are calculated at NLO with massive b-quarks (4FS) using the four-flavour NLO NNPDF3.0 PDF set [129]. Events are matched to the PYTHIA8 parton shower using the Powheg method.

Generator sample	Process	Flavour Scheme	Matching	Tune
		Nominal MC		
POWHEG + PYTHIA8	$t\bar{t}$ NLO	5F	Powheg h_{damp}=1.5 m_t	A14
MG5_aMC@NLO + PYTHIA8	$t\bar{t} + V/H$ NLO	5F	MC@NLO	A14
		MC as Systematic variations		
POWHEG + PYTHIA8 RadLo	$t\bar{t}$ NLO	5F	Powheg h_{damp}=1.5 m_t	A14Var3cDown
POWHEG + PYTHIA8 RadHi	$t\bar{t}$ NLO	5F	Powheg h_{damp}=3.0 m_t	A14Var3cUp
POWHEG + HERWIG7	$t\bar{t}$ NLO	5F	Powheg h_{damp}=1.5 m_t	H7UE
MG5_aMC@NLO + PYTHIA8	$t\bar{t}$ NLO	5F	MC@NLO	A14
SHERPA$t\bar{t}$	$t\bar{t}$ +0,1 parton at NLO +2,3,4 partons at LO	5F	MEPS@NLO	SHERPA
		MC predictions for Comparison only		
SHERPA	$t\bar{t}b\bar{b}$ NLO	4F	MC@NLO	SHERPA
POWHEG + PYTHIA8	$t\bar{t}b\bar{b}$ NLO	4F	Powheg $h_{\text{damp}}=H_T/2$	A14
POWHEL + PYTHIA8	$t\bar{t}b\bar{b}$ NLO	4F	Powheg $h_{\text{damp}}=H_T/2$	A14
POWHEL + PYTHIA8	$t\bar{t}b\bar{b}$ NLO	5F	Powheg $h_{\text{damp}}=H_T/2$	A14

Table 3.1.: Summary of MC setups used for modelling the signal processes for the data analysis and for comparisons to the measured cross-sections and differential distributions. All samples use the NNPDF3.0NLO PDF set with the exception of the two SHERPA samples which use NNPDF3.0NNLO. The different blocks indicate the corresponding usage of the samples in this analysis. For details see Section 3.4.

3.4.4. Background Modelling

The MC generators used to model the signal processes are discussed in detail in Section 3.4.1. The background processes are explained in Section 5.1. This section discusses the MC generators and PDF sets used to model the background processes. The nominal samples to model the $t\bar{t}+b$-jets signal and background processes and their cross section normalisations to the highest order perturbative predictions are summarised in Table 3.2.

Single top

Samples of Wt and s-channel single top quark backgrounds are generated with Powheg-Box 2.0 using the CT10 PDF set. Overlaps between the $t\bar{t}$ and Wt final states are removed. Electroweak t-channel single top quarks are generated using the Powheg-Box v1 generator which uses the 4FS for the NLO matrix elements calculations together with the fixed four-flavour PDF set CT10f4. For this process, the top quarks are decayed using MadSpin, preserving all spin correlations. All single top quark samples are interfaced to Pythia 6.428 with the Perugia 2012 underlying-event tune. The single top quark t- and s-channel samples are normalised to the approximate NNLO theoretical cross sections [130].

W/Z+jets

Events containing W or Z bosons with associated jets are simulated using the SHERPA2.2.1 [52] generator. Matrix elements are calculated for up to 2 partons at NLO and 4 partons at LO using the COMIX [127] and OPENLOOPS [128] matrix element generators and merged with the SHERPA parton shower [112] using the ME+PS@NLO prescription [126]. The NNPDF3.0NNLO PDF set is used in conjunction with dedicated parton shower tuning developed by the SHERPA authors. The $W/Z+$ jets events are normalised to NNLO cross sections, computed using FEWZ [131] with the MSTW2008NNLO PDF set.

Dibosons

The diboson samples are generated using SHERPA following the same approach as for the W/Z+jets sample but with up to one (ZZ) or zero (WW,WZ) additional partons at NLO and up to three additional partons at LO. These are normalised to their respective NLO cross sections calculated by the generator.

Process	Generator	Type	Normalisation
	Signal Processes		
$t\bar{t}$	POWHEG + PYTHIA8	NLO + PS	NNLO+NNLL
$t\bar{t}+V/H$	MG5_aMC@NLO +PYTHIA8	NLO + PS	NLO
	Background Processes		
Single top (Wt-channel)	POWHEG +PYTHIA6	NLO + PS	NNLO
Single top (s-/t-channel)	POWHEG +PYTHIA6	NLO + PS	NLO
W/Z+jets	SHERPA 2.2.1	$\leq 2j$ @ NLO, $\leq 4j$ @LO + PS	NNLO
Dibosons (ZZ)	SHERPA 2.2.1	$\leq 1j$ @ NLO, $\leq 3j$ @LO + PS	NLO
Dibosons (WW, WZ)	SHERPA 2.2.1	$0j$ @ NLO, $\leq 3j$ @LO + PS	NLO

Table 3.2.: Summary of signal ($t\bar{t}+t\bar{t}V+t\bar{t}H$) and background MC samples used in this analysis. For details see Sections 3.4.1 and 3.4.4.

3.5. Modelling Systematic Uncertainties

The choice of a particular $t\bar{t}$ MC generator (nominal) used in the signal modelling may induce a bias in the prediction. Therefore, modelling uncertainties are assigned to the measurement by comparing the nominal MC to the alternative generators. The comparison is performed one by one with an alternative generator by targeting one modelling component at a time to minimize the correlation among the modelling variations.

- **Hard-scatter generator:** An uncertainty due to the choice of hard-scatter generator is evaluated using an MC sample generated using Sherpa 2.2.
- **Parton shower:** An uncertainty due to the choice of the parton shower and fragmentation/hadronization model is evaluated using a MC sample with the same hard-scatter generator as the nominal MC sample, but using a different parton shower generator, HERWIG7, as opposed to PYTHIA 8.
- **Radiation:** An uncertainty on the amount of additional radiation present in $t\bar{t}$ events is evaluated by comparing the nominal POWHEG+PYTHIA 8 MC sample to two samples with parameters chosen to modify the amount of additional radiation. The two samples are both generated using the same versions of POWHEG and PYTHIA 8 as for the nominal $t\bar{t}$ MC sample but using the A14 VAR3C set of tunable parameters and different values of the h_{damp} parameter and the renormalisation and factorisation scales.

- **Parton distribution function:** The uncertainty due to the choice of PDF is evaluated following the PDF4LHC prescription [132] using event weights that are available in the nominal POWHEG + PYTHIA8 sample. The event weights allow the sample to be reweighted to produce predictions as if the MC sample was originally generated using that PDF set. The PDF4LHC uncertainty is evaluated by varying the input eigenvectors and then taking the standard deviation of variations as detailed in [132].

Chapter 4.

Object Reconstruction

The electronic signatures in different layers of the ATLAS detector are used to reconstruct and classify individual final state particles coming from pp collision. The analysis of $t\bar{t} + b$-jets process requires the reconstruction, identification and isolation of electrons, muons and jets as well as identification of b-jets. An event recorded in the ATLAS detector after the HLT online trigger is reconstructed through a sequence of *offline* algorithms. The same set of algorithms are also applied on the MC simulations for the reconstruction and classification of these physics objects. The simulated samples are then compared with data and calibrated in order to establish accurate modelling of the detector effects.

The procedure of object classification starts with the reconstruction of charged particle tracks and their association with the interaction point, called vertex, as discussed in Section 4.1. Section 4.2 and Section 4.3 provide the details about the reconstruction, identification, isolation and calibration of electrons and muons, respectively. The algorithms used for the reconstruction and calibration of jets are mentioned in Section 4.4. The jets originating from b-quarks need to be identified using a b-tagging algorithm explained in Section 4.5. Moreover, the systematic uncertainties associated with these measurements are also mentioned in each of these sections which contribute to this analysis. Section 4.6 briefs the missing transverse energy measurement which corresponds to the particles passing the detector undetected. A process called overlap removal is described in Section 4.7 which is employed to resolve any ambiguity between two objects reconstructed from the same information. These steps are generic for all analyses performed in the ATLAS collaboration and are therefore referred from [133] and [134].

Section 4.8 provides the definitions of objects at particle level which are applicable only to the simulated MC samples. To obtain the particle level information for data, a dedicated method, called unfolding, is required, which is explained in Chapter 6.

4.1. Tracks and Vertices

The magnetic field in the inner detector (ID) causes the charged particle to move in helical trajectories or tracks. The hits generated by a charged particle in different layers of the ID are combined to reconstruct a track. The set of parameters describing a track are (d_0, z_0, ϕ, θ and $q/|\vec{p}|$), where d_0 and z_0 are the track impact parameters[1] in the transverse and longitudinal directions respectively. ϕ and θ are the azimuthal and polar angles as mentioned in the previous chapter. $q/|\vec{p}|$ is the ratio of charge over momentum which defines the orientation and curvature of the helical trajectory.

The track reconstruction algorithms [135] use information from the ID. The *inside-out* algorithm starts by defining preliminary tracks called *seed tracks* from a set of three hits in the silicon detectors (pixel and SCT). A track is built by combining the seed track to the hits from the remaining layers of the silicon detectors moving towards the TRT. A hit entering in a track reconstruction must be compatible with the expected hit position in the detector predicted from the track parameters. The secondary charged particles coming from photon conversion or kaon decays might have very few or no hits in the silicon detectors to form a seed track. When the hits in the TRT are not associated with any seed track, a different approach is applied to reconstruct the tracks called *outside-in* algorithm. This algorithm starts the track seed in the TRT and trace back inwards to the SCT and pixel detectors.

The tracks are required to satisfy some quality cuts such as $p_{\mathrm{T}} > 400\,\mathrm{MeV}$, $|\eta| < 2.5$ and a minimum of seven hits in the silicon detectors [136]. These tracks are used to identify the charged particles and to measure their momenta. Additionally, the tracks are extrapolated to determine the interaction point where the hard scattering took place as explained below.

Once the tracks are reconstructed, a dedicated *vertex finder algorithm* [137, 138] is employed to reconstruct the vertices in an event. A vertex seed is created by finding the maximum of the impact parameter distribution of the reconstructed tracks. An iterative χ^2 based fitting is used on the vertex seed and the tracks around it which refits the seed position. The tracks that are incompatible with the seed vertex by more than 7σ are used to seed a new vertex. This procedure is employed iteratively until all the tracks are associated to a certain vertex. In this way the candidates for the primary vertex (PV) are defined. The vertex with the highest sum of p_T^2 of tracks is selected as the primary vertex, where the hard scattering interaction occurred. The remaining candidates of PVs are considered to be pileup interactions. Vertices incompatible with the beam collision region are considered as secondary vertices (Section 4.5).

[1] the transverse impact parameter, d_0, is the distance of closest approach of the extrapolated track to the interaction point and z_0 is the longitudinal impact parameter defined as the distance between the interaction point and the point on the track that determines d_0.

4.2. Electrons

Electrons are reconstructed in the central region of the ATLAS detector within $|\eta| < 2.47$ by matching the energy deposits in the electromagnetic calorimeter to tracks in the ID. An electron deposits its energy in many cells in three successive longitudinal ECal layers after the presampler. All these cells where an electron deposits its energy are collectively called *clusters*.

The first step of electron reconstruction [139, 140] is the construction of clusters. For this purpose, the ECal is divided into a grid of $N_\eta \times N_\phi = 200 \times 250$ elements in the $\eta - \phi$ space with size $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$, called towers. The energy deposited in the cells spanning several towers are summed in the longitudinal layers. The clustering algorithm [141] uses a window of 3×5 towers to search for a cluster seed with a minimum transverse energy E_T of 2.5 GeV. The next step is the association of a track with the cluster seeds. The tracks with $p_{\mathrm{T}} > 0.5\,\mathrm{GeV}$ are extrapolated from their last measured point in the tracking volume to points of impact in the middle layer of the ECal. A track and a cluster are considered to be matched when the distance between the track impact point and the seed cluster barycentre is within $|\Delta\eta| < 0.05$ and an electron candidate is reconstructed. To account for the bending of the tracks in the azimuthal plane due to bremsstrahlung losses, the size of track-cluster matching window is $\Delta\phi < 0.1$ [142]. If more than one track is matched with a cluster seed, the track with minimum distance from the seed position or with maximum number of hits in the silicon detector is selected. A cluster is classified as a photon candidate in absence of a matching track. Electrons are distinguished from converted photons (photon conversion into e^+e^- pair) by investigating the presence of pairs of close-by tracks originating from a vertex displaced from the interaction point. In Run-2, an additional requirement is imposed that the electron tracks are required to be compatible with the primary vertex in order to reduce the background from conversions and secondary particles. This is performed by following $|d_0|/\sigma_{d_0} < 5$ and $|z_0 \sin\theta| < 0.5\,\mathrm{mm}$ conditions. In the last step, the cluster sizes are optimised to take into account the overall energy distribution in the different regions of the ECal. The energy of the electron cluster is recomputed by enlarging its size to 3×7 towers in the barrel region and to 5×5 towers in the end-caps. The total energy of the reconstructed electron is obtained by adding contributions of: the cluster energy, the estimated energy deposits outside the cluster (lateral leakage), the estimated energy deposit in the material in front of the ECal and beyond the ECal (longitudinal leakage).

The reconstruction efficiency is defined as the ratio of number of reconstructed electron clusters matched to a track passing certain quality criteria to the total number of clusters with or without a matching track. A data-driven approach is used to determine the reconstruction efficiency using $Z \to ee$ events selected in data and MC as a function of E_T and η. It has a mild

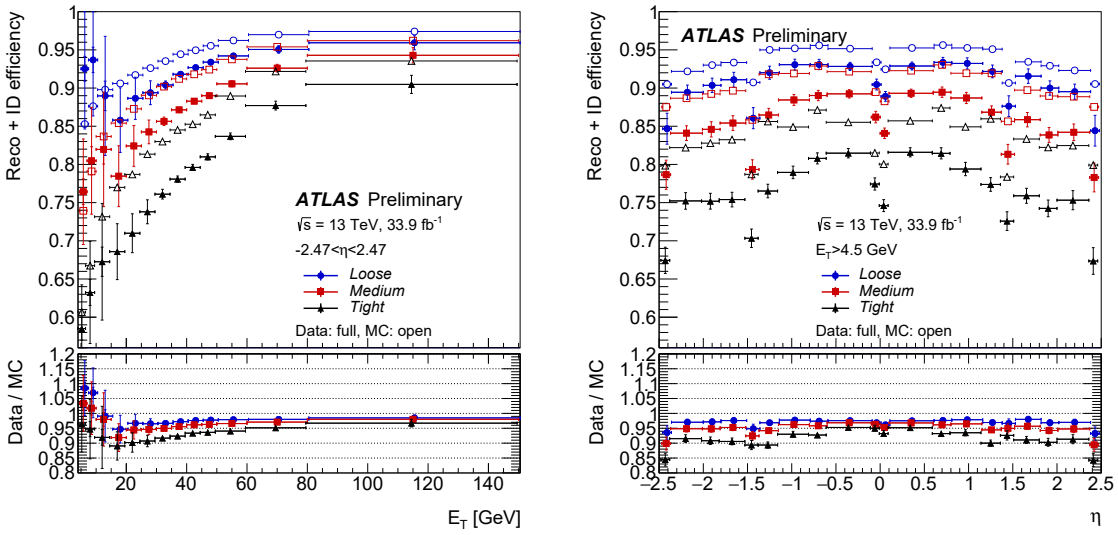


Figure 4.1.: Electron reconstruction and identification efficiencies in data and MC as a function of transverse energy E_T (left) and pseudo-rapidity η (right) of the electron. The uncertainty displayed includes both statistical and systematic uncertainties [143].

dependence on E_T, with values ranging from 98% for $\sim$ 20 GeV up to 99% for $E_T > 50$ GeV. The efficiency is around 98% in the central region of the detector except the transition region where it drops significantly [140]. Therefore, the electrons falling in the transition region are not considered in this analysis.

Not all the electrons obtained from the reconstruction step are necessarily the signal electrons i.e. electrons originating from heavy resonances like $W \to e\nu$ and $Z \to ee$ which are also called *prompt* electrons. Electrons coming from Dalitz decays, semi-leptonic decays of heavy quarks, photon conversion or hadrons misidentified as electrons are also obtained which are referred to as *non-prompt* electrons. An identification criterion [140] is implemented on the reconstructed electron candidates to reject these backgrounds as much as possible. It is based on a multivariate likelihood technique which utilizes discriminating variables such as the shape and size of the electromagnetic showers produced by the electrons in the ECal and track properties. Based on the requirements imposed on the likelihood discriminant, three levels of identification are defined : *Loose*, *Medium* and *Tight*, in increasing order of background rejection but decreasing signal efficiency. The *Tight* identification criterion is used in this analysis. The combined electron reconstruction and identification efficiencies in $Z \to ee$ events for $E_T > 15$ GeV and $J/\psi \to ee$ events for $7 < E_T < 20$ GeV are shown in Figure 4.1.

Electrons are required to be isolated, to further discriminate between signal and background, by quantifying the energy of the particles produced around the electron candidate. This allows the extraction of prompt electrons from non-isolated electron candidates. The isolation is done

both in ECal and ID using two discriminating variables namely calorimetric isolation and track isolation. The calorimetric isolation is defined by subtracting the electron cluster energy from the total deposited energy within a cone of radius $\Delta R = 0.2$ around the cluster. The track isolation variable is determined by subtracting the p_{T} of the track of the primary electron candidate (p_T^e) from the total p_{T} of all tracks in a cone of varying size $\Delta R = \min(10\,GeV/p_T^e, 0.2)$ around the candidate electron track. The selection on these isolation variables can be either constant or as a function of p_{T}. Different isolation operating points are defined by combining the measurements from these two discriminating variables. The *Gradient* isolation operating point is used in this analysis. The isolation efficiency is the ratio of the number of electrons passing certain isolation criteria to the total number of reconstructed and identified electrons. The isolation efficiency for the *Gradient* operating point is measured to be 90(99)% for p_{T} = 25(60) GeV from $Z \to ee$ events [140].

Finally, the calibration of the energy measurement of the electrons is performed [144]. The first aspect is simulation based calibration to recover the energy loss outside of the cluster and in passive material. A multivariate calibration is applied to both data and simulation. In addition, data-driven corrections are implemented to mitigate the non-uniformity in the detector response. Further, any residual disagreement between data and MC in the energy scale and resolution is accounted for by comparing the $Z \to ee$ mass distribution of data and MC. The uncertainty associated to the electron energy scale and resolution are 0.5% and 1%, respectively.

In addition, the MC simulations are corrected for the differences observed in the reconstruction, identification, isolation and trigger efficiencies of data and MC. Scale factor (SF) defined as the ratio of the efficiency of data over MC is applied on the MC. The SFs derived as a function of E_T and η, deviate only by few percent from unity. The uncertainty assigned to the SF is below 1% for $E_T > 25$ GeV [140].

4.3. Muons

Muon reconstruction is performed by combining the track information from the inner detector (ID) and the muon spectrometer (MS). The reconstruction of track segments from the ID is described in Section 4.1. Track reconstruction in the MS starts with the search for hit patterns in each muon chamber to form local track segments. The local track segments are then combined to form a MS track. The muon reconstruction algorithm [145] combines the ID-MS track segments based on the information provided by the ID, calorimeter and MS. Depending upon the subdetectors used in the reconstruction, four "muon types" are defined: combined muons, segment-tagged muons, calorimeter-tagged muons and extrapolated muons. In this

analysis combined muons are used. In this method the tracks are reconstructed independently in the ID and the MS, then a global re-fit is performed that uses the hits from both ID and MS to form a combined track.

Muon identification is performed to suppress the background mainly from pion and kaon decays. Four identification selections are provided for different physics analyses requirements, these are *Loose*, *Medium*, *Tight* and *High-*p_{T} [145]. Each selection provides a definite signal efficiency and background rejection. In this thesis, muons are required to pass the *Medium* identification criterion.

The muons are also required to be isolated using different isolation working points to further reduce the contamination of background muons from the semi-leptonic or hadronic decays. As for the electrons, muon isolation operating points also combine both the calorimeter and track variables. Muons are isolated using the *Gradient* working point in this analysis.

The efficiencies for the reconstruction, identification and isolation of the muons are measured both in data and MC using $Z \to \mu\mu$ events for $p_{\mathrm{T}} > 15\,\mathrm{GeV}$ and $J/\psi \to \mu\mu$ events for $5 < p_{\mathrm{T}} < 15\,\mathrm{GeV}$.

Momentum scale and resolution are also studied for muons as done for the electrons because even though the simulations contain accurate description of detector, the level of detail is not enough to describe these at per mill and percent level respectively. The agreement between data and simulation is achieved by applying a set of corrections to the simulated muon momentum. This is determined by the position and width of the Z peak in the $Z \to \mu^+\mu^-$ events [146]. The uncertainty on the muon momentum scale and resolution measurements are 2.3% and 0.05% respectively. The SF applied to the MC simulation to reproduce the measured data efficiencies has an uncertainty of below 1% for $p_{\mathrm{T}} > 25\,\mathrm{GeV}$.

4.4. Jets

Jets are the showers of particles formed by the fragmentation and hadronization of quarks and gluons, finally depositing their energy in the calorimeter systems. The process of jet reconstruction attempts to reconstruct the total energy and position of the particle shower to estimate the four-vector momentum representing the initial hard scatter parton. Jet reconstruction utilizes a clustering algorithm [147] which searches for topologically connected clusters of energy deposited cells in the electromagnetic and hadronic calorimeters. This algorithm is based on the significance of the energy deposited in a cell, E_{cell}, which is defined as the

energy deposited above the noise level. The noise level, σ_{cell}, is the sum in quadrature of the standard deviations of the distributions of electronic and estimated pile-up noise. A topological cluster or "*topo-cluster*" is seeded from a cell with $E_{\text{cell}} > 4\sigma_{\text{cell}}$. The topo-cluster is built in three-dimensional space by adding cells adjacent to the seed cell iteratively until no adjacent cell has significant energy greater than or equal to $2\sigma_{\text{cell}}$. The topo-cluster is completed by adding a single layer of adjacent cells with $E_{\text{cell}} > 0$. The procedure is applied on each of the seed cells to form topo-clusters.

Jet candidates are reconstructed by clustering the topo-clusters using the *anti-k_t* jet clustering algorithm [148]. One of the very important property of this algorithm is the *Infrared and Collinear* (IRC) *Safety*, which ensures that the shape of a jet is not influenced by soft radiation or collinear splitting. The algorithm iterates over every pair of topological clusters and aggregates them if they satisfy distance criteria. These distance parameters are defined as

$$d_{ij} = \min\left(\frac{1}{k_{t_i}^2}, \frac{1}{k_{t_j}^2}\right)\frac{\Delta R_{ij}^2}{R^2} \qquad \text{and} \qquad d_{iB} = \frac{1}{k_{t_i}^2} \tag{4.1}$$

and $\Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$, where $k_{t,i}, y_i, \phi_i$ are the transverse momentum, rapidity and azimuthal angle fo the topo-cluster i. d_{ij} is the distance between topo-clusters i and j, and d_{iB} is the distance of i from the beam-axis. The parameter R is the reference radius of the jet candidate and can be chosen freely, usually between [0.4, 1]. If d_{iB} is the smaller of the two distance parameters, then the topo-cluster i is considered as a complete jet and removed from the list of iterative algorithm process. Otherwise, if d_{ij} is smaller, the topo-clusters i and j are merged together and the list is updated for new iteration. This process continues until the list is empty which means all clusters are identified as jets. The value of R used in this analysis is 0.4.

The reconstructed jets undergo several steps of calibration [149] in order to correct for the energy scale of the jet reconstructed in the detector to correspond to the energy scale of the truth jet at particle level. This procedure accounts for the differences in responses of the two calorimeters, dead material in the detector, shower leakage and pile-up events. These corrections are derived from both MC simulation and data. The steps used to calibrate jet energy scale (JES) are applied in sequence as explained below.

- **Origin Correction:** This step recalculates the four-momentum of the jet such that it originates from the hard scattering primary-vertex rather than the geometrical centre of the ATLAS detector. The jet energy remains unaffected in this step, instead the η and ϕ resolution of the jet is improved.

- **Pile-up Correction:** This correction is applied to mitigate the effect of additional energy deposited from the in-time and out-of-time pile-up within the jet radius. Contribution to the jet energy from pile-up is estimated per-event from the product of the area of jet, A, and transverse momentum density ρ in the $\eta \times \phi$ plane. After the jet-area based correction, residual corrections are applied that are parametrized by factors α, β as

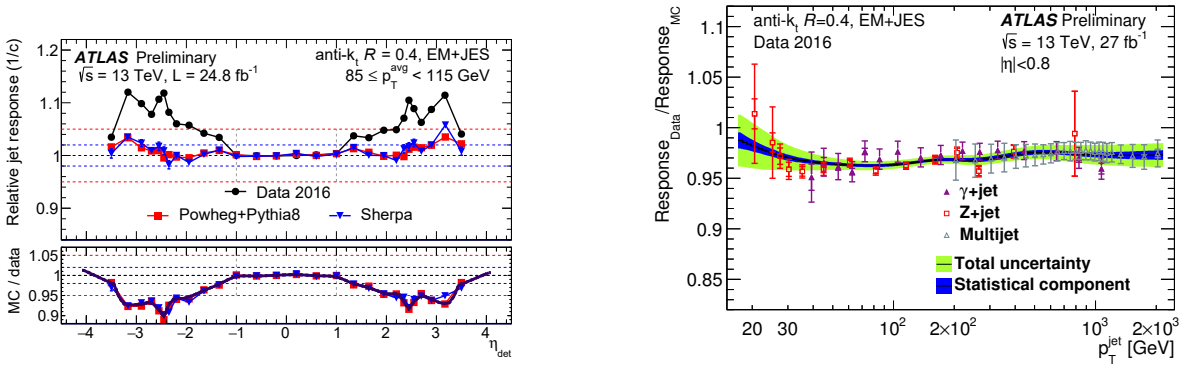
$$p_T^{\text{corr}} = p_T^{\text{reco}} - \rho \times A - \alpha \times (N_{\text{PV}} - 1) - \beta \times < \mu > \tag{4.2}$$

  where p_T^{corr} is the pile-up corrected p_{T} and p_T^{reco} refers to p_{T} of the reconstructed jet before any corrections applied. The last two terms in Equation 4.2 represent the residual corrections for the p_{T} dependence on the number of primary vertices, N_{PV}, and the average number of iterations per bunch crossing, $< \mu >$.

- **Jet energy scale and η calibration:** After the origin and pile-up correction, the next step is to match the reconstructed jet energy at the electromagnetic (EM) scale to the energy scale of the truth jet at particle level. The calibration is derived from MC simulation and a reconstructed colorimeter jet is considered to be matched to a truth jet if the distance between them is $\Delta R < 0.3$. To avoid the ambiguities between the calorimeter jets to truth jets, only isolated jets are used. A jet is isolated if there is no calorimeter jet of $p_{\text{T}} > 7\,\text{GeV}$ within $\Delta R = 0.6$ and only one jet of $p_{\text{T}} > 7\,\text{GeV}$ within $\Delta R = 1.0$. This corrects for the energy losses in the transition regions between calorimeters and in inactive material. Also, biases occur when a jet encompasses two calorimeters with different geometry and technology and crosses layers with different granularities. The η calibration is applied in order to make the jet response uniform in η.

- **Global Sequential Calibration :** The last correction derived using MC simulation is the global sequential correction. It accounts for the residual dependence of the JES on longitudinal and transverse features of the jets mainly due to the parton that initiated the jet. For example, a gluon-initiated jet is softer leading to lower calorimeter response and a wider transverse profile whereas a quark-initiated jet includes hadrons with higher fraction of p_{T} and penetrates further into the calorimeter. These corrections are applied based on the properties of tracks associated to jets, topology of energy deposits in the calorimeter and further in the muon spectrometer, if any. The mean of the jet energy is unchanged in this step.

- ***In situ* Calibration:** The last stage of the jet calibration accounts for differences in the jet response between data and MC simulation. Such differences arise from imperfect description of detector material and response in the MC simulation. The *in situ* corrections are derived from data by balancing the p_{T} of a jet against other well measured physics objects and are applied sequentially. First is the *η-intercalibration* which corrects for the p_{T}



(a) **(b)**

Figure 4.2.: (a) Result of η-intercalibration showing the relative jet response as a function of η in data and MC (top panel) and their ratio in the lower panel. Data is calibrated to MC and the resulting residual correction is shown as the thick magenta line in the lower panel. (b) Jet response ratio of the data to MC as a function of p_{T} for Z/γ+jet and multijet balance. Black line inside blue band is the combined result of *in situ* calibration [150].

of forward jets with $0.8 < |\eta| < 4.5$ by using the well measured jets in the central region, $|\eta| < 0.8$ in dijet events, thereby flattening the response across the detector (Figure 4.2a). This is followed by Z/γ+jet calibrations for jets up to 950 GeV with $|\eta| < 0.8$. The last step is the *multijet balance* (MJB) to extend the calibration to p_{T} of 2 TeV. A high-p_{T} jet is balanced by a recoil system composed of several lower-p_{T} jets where the low p_{T} jets are already calibrated. Finally, the Z/γ+jet and MJB calibration factors are combined as shown in Figure 4.2b.

In addition to the uniform correction in p_{T} of the jets by subtracting the additional energy deposits from the pile-up effect, removal of pile-up jets is also applied to correct for local fluctuations in pile-up activity. To achieve this, *Jet Vertex Tagger* (JVT) [151] is implemented for identifying jets that originate from hard scattering interaction. It is a multivariate combination of two variables: corrJVF and $R_{p_{\mathrm{T}}}$ which are related to the ratio of the sum of p_{T} of all tracks associated to a jet that are matched to the hard scatter primary vertex to the p_{T} sum of all associated tracks. The output of the two-dimensional likelihood JVT discriminant lies between 0 and 1 where jets from pile-up peak at 0 and jets originating from hard scattering at 1. In order to suppress pileup jets, a requirement of JVT > 0.59 for central jets within $|\eta| < 2.4$ and $p_{\mathrm{T}} < 60$ GeV is applied that provides 92% efficiency in selecting hard scattered jets.

Jet energy measurement uncertainty

The uncertainties associated with the jet selection arise from the calibration and efficiency of Jet Energy Resolution (JER), JES and JVT. The JES uncertainty consists of several components which can be classified as detector, modelling, statistical, η intercalibration, high p_{T} jet

response, jet flavour, pile-up and punch through. The full set of components are detailed in Reference [149]. In this analysis JES uncertainties are evaluated using a set of 21 nuisance parameters. The break-down of the nuisance parameters is as follows:

- Eight nuisance parameters from the in-situ calibrations obtained from an initial set of 75 nuisance parameters by taking the seven principal components of the greatest magnitude and combining the rest in quadrature. The loss in correlations due to this procedure is at the percent level.
- Four pileup uncertainties to account for potential mis-modelling of $N_{\mathrm{PV}}, \mu, \rho$ and the residual p_{T} dependence.
- Three η-intercalibration uncertainties taking into account physics modelling, statistical uncertainties and the method non-closure in the $2.0 < |\eta_{\mathrm{det}}| < 2.6$ region.
- Three uncertainties to account for differences in the jet response and the simulated jet composition of light-quark, b-quark and gluon-initiated jets.
- One uncertainty due to punch-through.
- One high-p_{T} jet uncertainty, applied to jets with $p_{\mathrm{T}} > 2\,\mathrm{TeV}$, which can not be covered by the in-situ analyses.
- One uncertainty to account for the non-closure of the absolute JES calibration of fast-simulation jets.

The total JES uncertainty coming from all the nuisance parameters ranges from 6% for jet p_{T} > 25 GeV to less than 2% for p_{T} > 100 GeV. The total uncertainty due to the JES is one of the dominant uncertainties in this analysis.

The JER is measured using both data and MC. First, the truth resolution is measured by comparing the truth and reconstructed jet p_{T} in MC as a function of the jet p_{T}. Secondly, an in-situ measurement of the JER is made and used to compare the resolution in data and MC [152]. The jet energy in simulated samples is smeared by a Gaussian with width dependent on p_{T} and η of the jet to match the jet energy resolution in data. The systematic uncertainty assigned to JER is less than 4% for $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ [153, 154].

The JVT is calibrated using $Z \to \mu\mu$ +jets events. Scale factors binned in jet p_{T} are applied to each event to correct for small differences in the JVT efficiency between data and MC. An uncertainty of less than 1% is associated with the SF.

4.5. *b*-tagging

The process of identifying a jet containing a b-hadron is referred to as b-tagging. It is very crucial to identify b-jets and separate them from c-jets and light-jets (coming from u, d, s quarks or gluons) for the analyses containing b-quarks in the final-state like the one presented here. Some distinctive properties of b-hadrons such as the long lifetime ($\tau \sim 1.5\,\text{ps}$, $c\tau \sim 450\,\mu\text{m}$), higher mass and decay multiplicities are exploited in order to identify and separate b-jets. In the energy scale of $p_{\text{T}} > 20\,\text{GeV}$, the b-hadrons travel on average several millimetres before they decay, leading to topologies with secondary vertices displaced from the primary vertex of the hard-scattering event. Also, due to the long lifetime of b-hadrons, tracks generated from the decay products of b-hadrons tend to have large impact parameters as compared to the tracks originating from the primary vertex. Various algorithms have been developed by the ATLAS collaboration to identify b-jets utilizing these characteristics.

The b-tagging algorithm used in this analysis is called MV2c10 [155] which is a multivariate algorithm which combines the information from several other algorithms described below:

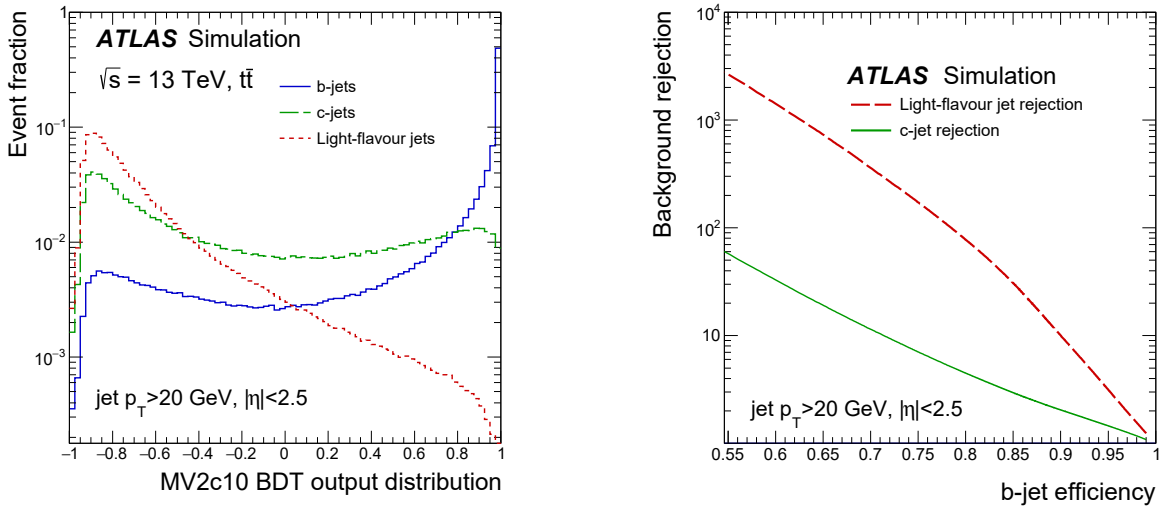
- **The Impact Parameter based algorithms** [156] are based on the significance of longitudinal and transverse impact parameters expressed as $z_0 \sin\theta / \sigma_{z_0 \sin\theta}$ and d_0 / σ_{d_0} respectively. These parameters are used to construct probability density function from MC simulation. At last, a log-likelihood ratio is computed to discriminate between different jet flavours.

- **Secondary vertex finder algorithm, SV1** [156] explicitly reconstructs a displaced secondary vertex within the jet.

- **The JetFitter** [157] is a decay chain multi-vertex reconstruction algorithm that utilizes the topology of b- and c-hadron decays inside a jet to reconstruct the full b-hadron decay chain. A Kalman-filter [158] is used in order to determine a common line on which the primary vertex and the b- and c-hadron decay vertices lie, approximating the b-hadron flight trajectory.

The MV2c10 algorithm is the final output which takes as input all the above mentioned algorithms in a Boosted Decision Tree (BDT). The BDT is trained on a $t\bar{t}$ sample with b-jets as signal and a mixture of 7% c-jets and 93% light-flavoured jets as background. The output of BDT shown in Figure 4.3a is referred to as MV2c10 score or b-tagging discriminant value.

The performance of a b-tagging algorithm is quantified as b-tagging efficiency which is defined as the efficiency to identify a jet as a b-jet. There are also possibilities of tagging a jet containing a c-hadron or a jet originating from light-flavoured parton as a b-jet because of the limitations of the b-tagging algorithm and inefficiencies in the tracks, vertices and jets reconstructions.



(a) **(b)**

Figure 4.3.: (a) The MV2c10 output for b-jets (solid line), c-jets (dashed line) and light-flavour jets (dotted line). (b) The light-flavour jet (dashed line) and c-jet (solid line) rejection factors as a function of the b-jet tagging efficiency of the MV2c10 b-tagging algorithm. These performances are evaluated using simulated $t\bar{t}$ events [159].

The inverse of the efficiency for tagging a light-flavour jet or a c-jet as a b-jet is referred to as the rejection rates for light-flavour jets and c-jets, respectively. Figure 4.3b shows the corresponding c-jet and light-jet rejection factors as a function of b-jet efficiency. The MV2c10 BDT score lies between 0 and 1; by imposing a cut on the MV2c10 score a specific b-tagging efficiency with the corresponding purity can be selected. Four operating/working points (WP) are defined based on the MV2c10 score for specific b-tagging efficiencies of 60%, 70%, 77% and 85% and corresponding rejection rates as listed in Tables 4.1. It should be noted that increasing the b-jet efficiency decreases the c-jet and light-jet rejection rates. For instance, the 60% b-tagging efficiency WP has the highest purity whereas the 85% WP has the lowest purity.

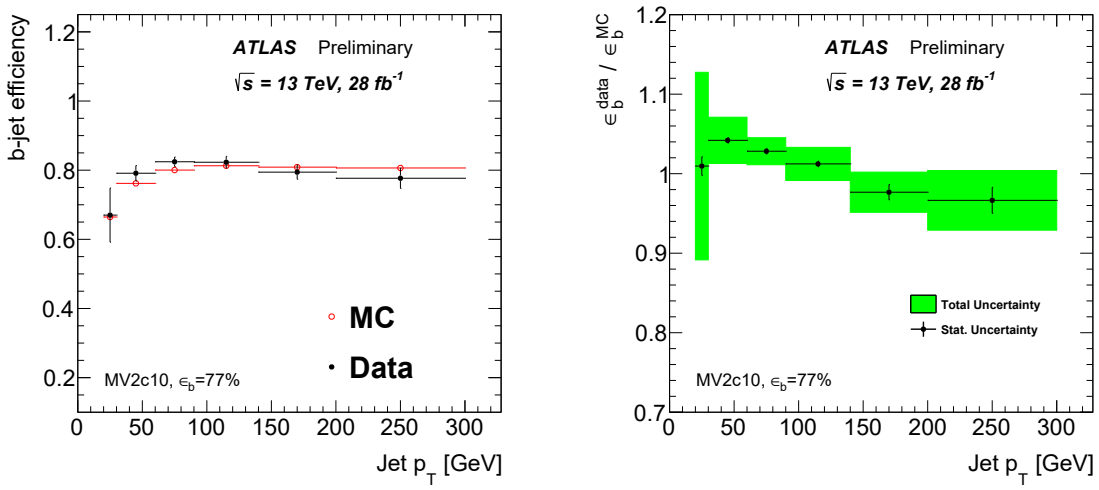
BDT Cut Value	b-jet Efficiency in %	c-jet Rejection	Light-jet Rejection
0.9349	60	34	1538
0.8244	70	12	381
0.6459	77	6	134
0.1758	85	3.1	33

Table 4.1.: Operating points for the MV2c10 b-tagging algorithm, including benchmark numbers for the efficiency and rejections rates. These values have been extracted from $t\bar{t}$ events with jet p_{T} above 20 GeV [155].

One straight forward way to apply b-tagging is to choose one of the working points with a desired b-jet efficiency and select the jets above that cut value. In this approach a jet is

classified either tagged or not at a given working point. A more sophisticated approach, called *pseudo-continuous* (PC) b-tagging, is available in which the MV2c10 distribution is divided exclusively in five bins such that each bin corresponds to a certain range of b-tagging efficiencies. The bin range from 1-5 corresponds to efficiencies of 100%-85%, 85%-77%, 77%-70%, 70%-60%, and <60% respectively. In this case, the jets can be classified in five divisions according to the purest and most efficient b-jet selection. For example, if an 85% WP is chosen, a jet can further be classified in four divisions based on the purity and efficiency instead of two classes only. The pseudo continuous b-tagging method is used in this analysis with a working point of 77% b-tagging efficiency.

The performances mentioned above are derived from MC simulation, therefore *b-tagging calibration* [160] is applied to correct for the modelling effects. Scale factor, defined as $SF = \varepsilon_{\text{data}}/\varepsilon_{\text{MC}}$, is applied to correct the performance in MC to data. $\varepsilon_{\text{data}}$ and ε_{MC} are the efficiencies measured in data and MC respectively. The efficiency in data is measured from a data sample which is very pure in the particular jet flavour. The b-tagging efficiency in data is determined from the di-leptonic $t\bar{t}$ sample and the scale factors are derived as a function of jet p_{T} and η. Figure 4.4 shows data and MC b-jet efficiencies and SFs for the 77% WP as a function of jet p_{T}.



(a) **(b)**

Figure 4.4.: b-tagging efficiency for data and MC (left) and the scale factor computed from their ratio (right) for the MV2c10 b-tagging algorithm at 77% working point as a function of jet p_{T}. Statistical errors (error bars) and total errors (shaded region) are shown [161].

As the b-jet tagging efficiency is calibrated using dilepton $t\bar{t}$ events, the c-jet mistag rate is calibrated using semi-leptonic $t\bar{t}$ events [155, 162]. While for light-jets a negative tag method is used to calibrate the mis-tagged rate [163]. The uncertainties from these calibrations are propagated into this analysis using an eigenvector method. In total there are:

- 30 uncertainties associated with the calibration of b-jets
- 15 uncertainties associated with the calibration of c-jets
- 80 uncertainties associated with the calibration of light-jets.

The uncertainties are computed by varying the SFs by $\pm 1\sigma$ and measured with a precision of 2% for $p_{\mathrm{T}} > 60\,\mathrm{GeV}$ whereas for low p_{T} it is worse as 6%. Due to the large number of b-tagged jets used in this analysis (at least three), the total uncertainty due to b-tagging is one of the dominant uncertainties in this analysis.

4.6. Missing Transverse Energy

The partons colliding at LHC carry no initial transverse momentum, therefore by momentum conservation, the sum of p_{T} of the collision products should also be zero. Any inequality in the sum is measured as *Missing Transverse Energy* (MET) or $E_{\mathrm{T}}^{\mathrm{miss}}$. The missing energy indicates the presence of particles which traverses the detector without interacting. In the case of the Standard Model, these are neutrinos. Objects lying outside the detector acceptance, those reconstructed poorly or not reconstructed at all also contribute to $E_{\mathrm{T}}^{\mathrm{miss}}$.

The $E_{\mathrm{T}}^{\mathrm{miss}}$ is evaluated by calculating the negative vector sum of the momenta of all reconstructed hard-scattered objects. An additional correction from calorimeter clusters or tracks from primary vertex that are not matched with any of the reconstructed objects (soft term) [161] is also included. The x and y components of $E_{\mathrm{T}}^{\mathrm{miss}}$ is then given by

$$E_{x(y)}^{\mathrm{miss}} = -E_{x(y)}^{e} - E_{x(y)}^{\gamma} - E_{x(y)}^{\tau} - E_{x(y)}^{\mathrm{jets}} - E_{x(y)}^{\mu} - E_{x(y)}^{\mathrm{soft}} \tag{4.3}$$

All objects (electron, photon, tau, jets and muons) that contribute to the missing transverse energy are calibrated in order to obtain the correct magnitude of this measure.. The total $E_{\mathrm{T}}^{\mathrm{miss}}$ is then given by

$$E_{T}^{\mathrm{miss}} = \sqrt{(E_{x}^{\mathrm{miss}})^2 + (E_{y}^{\mathrm{miss}})^2} \tag{4.4}$$

4.7. Overlap Removal

The independent reconstruction and identification of electrons, muons and jets are performed based on the information from tracks and calorimeter clusters as explained above. Ambiguities can occur if two different object candidates are reconstructed from the same information which

leads to energy double counting or mis-reconstruction of an object. To resolve this issue an overlap removal procedure is employed that solves conflicts of overlapping candidates and classifies them as a single object.

If an electron and a muon share a track within $\Delta R\,(e,\mu) < 0.02$, the muon candidate is given preference over the electron candidate if it satisfies all the muon selection criteria. Since, the jet clustering algorithms incorporate the electromagnetic clusters in the clustering procedure, if an electron and a jet lie within $\Delta R\,(e,j) < 0.2$, the jet is removed in order to avoid counting the electron energy deposits twice. On the other hand, if there are remaining jets within $\Delta R\,(e,j) < 0.4$ then the electron is removed in order to reduce the impact of non-prompt leptons coming from the semi-leptonic decays of the jet constituents. If a jet is within $\Delta R\,(\mu,j) < 0.4$ of a muon, then the jet is removed if it has less than three tracks associated with it whereas if the jet has at least 3 tracks the muon is removed and the jet is kept.

4.8. Particle Level Object Definitions

As mentioned in Section 3.3, the outputs of the MC simulations are recorded at different stages and are referred to as detector and particle level. To compare a measurement with other experiments or with different MC models, particle level objects are defined. These objects closely match the definitions of the detector-level objects, defined in previous sections of this chapter, in order to minimise corrections from detector to particle level.

Particle-level objects are constructed from stable particles with a proper lifetime of more than 3×10^{-11} s. The definitions of the physics objects are as follows:

- **Electrons and muons:** Electrons and muons in the MC event record are required not to come from hadron decays[2] to ensure that the leptons are originating from real W or Z-boson. The direct matching of leptons with the W-boson is not required because the simulated $t\bar{t}$ sample is used to define the particle level. The collimated final state photon radiation of the leptons is hard to separate in detector level object reconstruction because of the insufficient detector resolution. To reproduce this effect at particle level, the four-vector momenta of all the stable photons in a cone of size $\Delta R = 0.1$ around the lepton direction are added to the electrons and muons. These *dressed leptons* are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Experimental details of the detector are not included in particle level definition, e.g. the transition region of $1.37 < |\eta| < 1.52$ is not rejected.

[2]Electrons and muons from tau decays are thus included.

- **Jets:** Particle level jets are clustered using all stable particles except those used in the definition of dressed electrons and muons above and neutrinos not from hadron decays, using the anti-k_t algorithm with a radius parameter of 0.4. The decay products of hadronically decaying τ leptons are therefore included. These jets do not include particles from pile up events but do include those from the underlying events. Decay neutrinos are included in the b-jet definition on particle-level to be independent of the decay model. Jets are required to have $p_{\mathrm{T}} > 25$ GeV and $|\eta| < 2.5$.

- **Jet flavour identification:** For the particle level, jets are identified as b-jets by requiring that at least one b-hadron with $p_{\mathrm{T}} > 5$ GeV is *ghost-associated* to the jet [164]. Ghost association is a procedure to match particles with jets, based on the "jet catchment area" [164]. In this method a single infinitely soft "ghost" particle is added to the particle which is clustered into a jet. In other words, the momentum of the particle is scaled by a negligible value and included in jet clustering using anti-k_t which is an infra-red and collinear safe algorithm. The infinitely small value of the added momentum ensures that the result of jet clustering is not altered. This method provides very reliable results for the closely lying or collinear jets which is very important for the final states like $t\bar{t}b\bar{b}$.

  A similar procedure is followed to define c-jets, with the b-jet definition taking precedence i.e. a jet containing one b-hadron and one c-hadron is defined as a b-jet. Therefore the jets containing a c-hadron but no b-hadron is defined as a c-jet. Jets that do not contain either a b- or c-hadron are considered to be light jets. Accessing this information correctly is very crucial for some truth level studies presented in Section 5.5.

- **Overlap between objects:** Electrons and muons that pass the selection criteria defined above are required to be separated from selected jets with $\Delta R(\ell, j) > 0.4$. This ensures consistency with the reconstruction level objects.

Chapter 5.

Event Selection and Background Determination

The previous chapters discussed the details of the theoretical basis and the experimental setup utilized in this analysis. The next step is the proper choice of analysis techniques which are derived from the characteristics of $pp \rightarrow t\bar{t} + b$-jets process in order to perform the best possible measurement.

Among different possible decay channels of $t\bar{t}$, the selection of $e\mu$ channel and the number of additional b-jets is explained in Section 5.1. This section also introduces the background processes which contaminate the process of interest. A dedicated event selection as explained in Section 5.2 is applied to enrich the signal and reduce the background contributions. These selected events still contain some remaining backgrounds which need to be determined precisely using relevant procedures discussed in Section 5.3. The final number of events for the signal and each background is presented in Section 5.4. The measurement of $t\bar{t} + b$-jets production depends on the determination of the background from other $t\bar{t}$ processes i.e. $t\bar{t}c$ and $t\bar{t}l$. Section 5.5 describes the data-driven template fit procedure used to extract $t\bar{t}b$ signal yields and estimate the $t\bar{t}c$ and $t\bar{t}l$ backgrounds. Section 5.6 lists all the sensitive observables used in this analysis and Section 5.7 shows the data-MC comparison for these observables at detector-level.

5.1. Signal and Background

Signal refers to the process of interest which is $pp \rightarrow t\bar{t} + b$-jets in this analysis. There are other physics processes which has the same final state as the signal but this final states is not the decay product of the signal. These processes contaminate the signal region and are called *backgrounds*. Experimentally, only reconstructed final state particles are measured, therefore it is difficult to separate their origin and to distinguish the signal and background processes.

The following will explain the signature selected for the signal and the possible background processes which enter in this analysis.

Signal processes

The $t\bar{t}$ decays are typically classified based on the decay of the two W bosons (Section 1.3.2). One of the decay channels of $t\bar{t}$ is classified as the di-leptonic channel where both W bosons decay leptonically. This analysis is performed particularly in one of the three possible di-leptonic channels referred to as the $e\mu$ channel where one W boson directly decays into a muon and a muon neutrino, and the other to an electron and an electron neutrino. The same flavour di-lepton decay channels (ee or $\mu\mu$), where both the W bosons decay either to two electrons or to two muons are not included in this analysis. This is to reduce the background contribution from the Drell-Yan process as discussed later. The W bosons can also decay into τ leptons ($WW \rightarrow \tau\tau$) in the di-leptonic channel. The τ lepton further decays either leptonically or hadronically. If the final state resulting from τ decays give the same signature as $e\mu$, it is also considered as part of the signal. Conclusively, one electron and one muon produced either directly in the decay of the W boson or via an intermediate τ lepton are included in the signal. Along with the two W bosons, the decay of the top quark pair results in two b-quarks as well. In this thesis, the *baseline selection* is a final state with an $e\mu$ pair and at least two b-quarks. It is referred to either as $t\bar{t}$ process or $\geq 2b$ in terms of the number of b-jets.

Despite having very low branching ratio, the di-leptonic channel is selected because it suffers from lower background contamination as compared to fully-hadronic and semi-leptonic channels. Since the pp collision at the LHC produces plethora of jets, it is difficult to distinguish the fully-hadronic and semi-leptonic signal signatures from the background processes. As mentioned in Section 4.5, the b-tagging algorithm sometimes tag a c- or light-jet as a b-jet. Since this analysis concerns the $t\bar{t} + b$-jets process, it suffers from inseparable $t\bar{t} + c$-jets and $t\bar{t}$ + light - jets background as well. This can be minimised by ignoring the hadronic decay of W boson into a cs pair, which has a high probability of being mis-tagged as b-jets.

The production of additional b-jets with the $t\bar{t}$ pair is of great importance for the test of QCD predictions. Referring back to Section 1.4, these additional b-jets can arise from QCD radiation with gluon splitting into $b\bar{b}$ or subsequent decay of associated bosons (H, W, Z). In this analysis, the origin of b-jets is not identified. Hence, the measurement is performed without separating the b-jets from QCD radiation or associated boson decays. Also, no distinction is made between the b-jets coming directly from the $t\bar{t}$ decay or additional b-jet sources. This is to avoid using simulation-based information to attribute b-jets to a particular production process which would lead to significant modelling uncertainties. In conclusion, the signal process contains a top quark pair with its subsequent decays and at least two additional b-jets leading

to the reference name $t\bar{t}b\bar{b}$ or $\geq 4b$ in terms of the number of b-jets as shown in Figure 5.1.

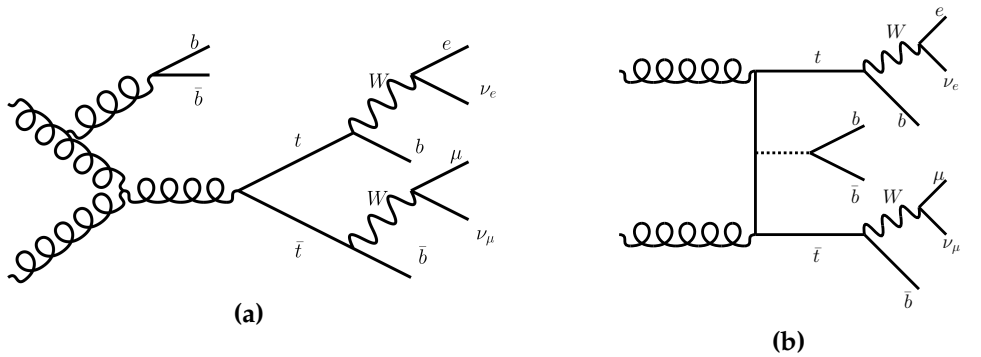


Figure 5.1.: Feynman diagrams presenting the full $t\bar{t}$ process with additional b-jets coming from QCD radiation (left) and from an associated boson decay (right).

The cross-section is also measured for the $t\bar{t}$ process with one additional b-jet, referred to as $t\bar{t}b$ or $\geq 3b$ in terms of the number of b-jets. It can be deduced from the previous paragraph that the additional b-quarks are produced in pairs but there are several possibilities of getting only one additional b-jet in the final state, experimentally. Sometimes one b-jet escapes the detector coverage or only one b-jet passes the event selection criteria which will be discussed in Section 5.2. Another possibility is getting events where the two b-quark showers are so collimated that these are identified as a single b-jet. These are some of the reasons behind having more events in the $t\bar{t}b$ case as compared to the $t\bar{t}b\bar{b}$ process. While measuring the cross-section for the $\geq 4\,b$-jets phase space is of great relevance to probe the $t\bar{t}H$ measurements, the region with $3\,b$-jets provides complementary information on the collinear $g \to b\bar{b}$ rate. Therefore, this measurement allows testing and putting constraints on the MC generators along with providing information for constraining the total $t\bar{t}b\bar{b}$ rate for the $t\bar{t}H$ measurements.

To summarize, in reconstructed pp-collision events, the signal is searched for in a dataset consisting of a reconstructed electron, reconstructed muon and two or more jets which are b-tagged. The measurements are performed with the $\geq 2b$, $\geq 3b$ and $\geq 4b$ selections and named as $t\bar{t}$, $t\bar{t}b$ and $t\bar{t}b\bar{b}$ processes respectively.

Background processes

The background processes contaminating the signal can be categorized into two types:

- those with two real prompt leptons from W or Z decays (including production via τ decays)
- those where at least one of the reconstructed lepton candidates is misidentified; categorised as non-prompt or fake leptons

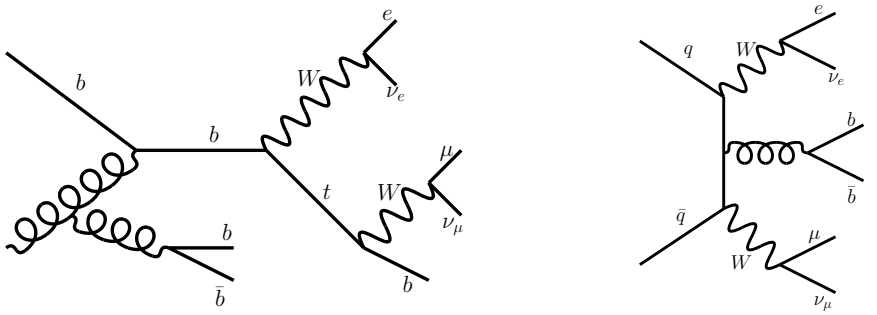


Figure 5.2.: Examples of some background processes leading to the same final state as the signal process. NLO Wt process (left) and WW diboson process (right) with associated additional b-jets.

The first category with two prompt leptons is dominated by single top production in the Wt channel. It also includes a small fraction of Drell-Yan and diboson (WW, WZ and ZZ) backgrounds. The second category contains events with at least one misidentified lepton in the selected sample such as a non-prompt lepton from the decay of a bottom or charm hadron, an electron from a photon conversion, or a muon produced from an in-flight decay of a pion or kaon, or hadronic jet activity faking a lepton.

Single Top production constitutes the dominant background in this analysis. The production of a single top quark in association with a real W boson is referred to as single top production in the Wt-channel or simply the Wt process (Figure 5.2a). The top quark decays further into a W boson and a b-quark. The subsequent decay of the two W bosons into an $e\mu$ pair and the production of b-jets from gluon splitting represents the same final state as the signal process and therefore this process enters the analysis selection.

Diboson processes namely WW, WZ and ZZ are the smallest background contribution in this analysis. The leptonic decay of both W bosons resulting in an $e\mu$ final state in the WW process (Figure 5.2b) can enter the analysis selection. In case of WZ and ZZ productions, the W boson decays to a lepton and a corresponding neutrino whereas the Z boson decays to a same-flavour lepton pair. Therefore, the WZ and ZZ processes also have one electron and one muon but with the presence of more leptons. The requirement of exactly one e and one μ reduces this background. In addition, the requirement of b-jets suppresses this background highly.

The **Drell-Yan** process also constitutes a small fraction of background. In the Drell-Yan process, a Z boson or virtual photon decays to two same flavour leptons $Z/\gamma* \rightarrow ee/\mu\mu/\tau\tau$. The event selection requirement of leptons of different flavour already reduces the background coming from the ee and $\mu\mu$ channels. However, the events originating from $Z/\gamma^* \rightarrow \tau\tau \rightarrow e\mu$ decays, accompanied by two or more b-jets, remain in the selection.

Non-prompt and fake leptons are the second category of backgrounds. Despite the rigorous algorithms implemented for the reconstruction, identification and isolation of leptons, elaborated in Chapter 4, misidentification of reconstructed objects may still occur. There are two possibilities for this to happen: an object is reconstructed as a lepton but without the presence of a real lepton called fake leptons and selection of a real lepton that does not originate from the heavy resonances (W or Z) called non-prompt leptons in this analysis. The *non-prompt leptons* are the major contributor in this category of backgrounds. The converted photons from π^0 decays to $\gamma\gamma$ are hard to distinguish from prompt electrons. A muon can be produced from the in-flight disintegration of charged mesons such as pions and kaons ($\pi^\pm \to \mu^\pm \nu_\mu$ or $K^\pm \to \mu^\pm \nu_\mu$) leading to non-prompt muon production. These contributions can lead to the signature of a prompt lepton. The *fake leptons* receive contribution, yet a very small one, from the W +jets processes. In this process, the leptonic decay of the W boson produces one real lepton and one of the associated hadronic jet can be mis-identified as the second lepton of different flavour. Therefore, it can enter the signal region with one real and one fake lepton resulting in an $e\mu$ pair and associated b-jets.

5.2. Event Selection

A set of selection criteria is applied on the data and simulated samples to suppress the background contamination in the signal region while retaining a high signal selection efficiency. The event selection is applied in multiple steps and the choice of selection cuts is motivated by the final-state and its kinematic properties of the signal process.

Trigger Selection

In real events (data), the objects are stored only after passing both L1 and HLT online trigger criteria as mentioned in Section 2.2.5. The same triggers are also implemented for the simulated events to obtain coherency between data and MC samples. In this analysis, the events are selected using single-electron or single-muon triggers. Table 5.1 summarizes the triggers imposed on data and MC samples in this analysis for the 2015 and 2016 data taking with the corresponding p_{T} thresholds.

For the electrons, a logical OR of three single-electron triggers are used. The 2015 trigger, `HLT_e24_lhmedium_L1EM20VH` requires an electron candidate passing the HLT p_{T} threshold of 24 GeV and the medium likelihood-based identification criterion seeded from L1 trigger with a p_{T} threshold of 20 GeV. The other two triggers correspond to a higher HLT threshold of 60 and 120 GeV satisfying medium and loose identification requirements respectively. In 2016,

Object	Trigger name	p_T (GeV)
	2015	
Electrons	HLT_e24_lhmedium_L1EM20VH	24
	HLT_e60_lhmedium	60
	HLT_e120_loose	120
Muons	HLT_mu20_iloose_L1MU15	20
	HLT_mu50	50
	2016	
Electrons	HLT_e26_lhtight_nod0_ivarloose	26
	HLT_e60_lhmedium_nod0	60
	HLT_e140_lhloose_nod0	140
Muons	HLT_mu26_ivarmedium	26
	HLT_mu50	50

Table 5.1.: The triggers used to select electrons and muons for the dataset collected in the year 2015 and 2016 and for the corresponding MC datasets.

the trigger with lower p_T threshold has in addition an isolation requirement to reduce the high trigger rate of leptons produced in hadron decays. "nod0" implies the absence of track impact parameter requirements.

For muons, a logical OR between two single-muon triggers is used. One corresponds to a threshold of 20 (26) GeV for 2015 (2016) dataset, satisfying loose (medium) isolation criteria and the other corresponds to a threshold of 50 GeV. The details on lepton identification and isolation criteria can be found in Sections 4.2 and 4.3.

Object Selection

Followed by the triggers applied to the data and MC events, the detector level objects are required to satisfy the kinematic selections. The event selection requires to have exactly two leptons with different flavour and opposite-sign (OS)[1], resulting in an $e\mu$ pair. At least one of the selected leptons needs to be matched to an electron or a muon trigger as mentioned above. Furthermore, the events must contain at least two jets of which at least two are b-tagged at the 77% efficiency b-tagging working point.

Electrons are required to have $p_T > 27$ GeV and $|\eta| < 2.47$. Candidates that fall in the transition

[1] Leptons with opposite sign of the electric charge are referred to as opposite-sign (OS) i.e. $e^-\mu^+$ or $e^+\mu^-$. Lepton pairs with same-sign are referred to as SS

region between the barrel and end-cap ($1.37 < |\eta| < 1.52$) are poorly measured and therefore not considered in this analysis. Electrons must pass *Tight* likelihood identification criteria and the *Gradient* isolation operating point. Muon candidates are required to have $p_{\mathrm{T}} > 27\,\mathrm{GeV}$ and $|\eta| < 2.5$. Muons must pass the *Medium* identification criterion and are isolated using the *Gradient* isolation.

Jets are required to have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. Jets with $p_{\mathrm{T}} < 60\,\mathrm{GeV}$ and $|\eta| < 2.4$ are required to be identified as originating from the primary vertex using a JVT cut value of 0.59. A jet cleaning requirement called LooseBadJet cleaning is applied to reject events where any jet originates from detector effects (like detector noise) or non-collision sources. The b-jets are identified using the MV2c10 tagger at 77% working point.

The procedure called overlap removal is applied to resolve any ambiguities of a single detector response being assigned to two objects by the reconstruction algorithms. Table 5.2 summarizes the kinematic requirements applied to the reconstructed objects and event selection for the baseline $t\bar{t}$ process. This baseline selection with at least 2 b-tagged jets is 95% pure in $t\bar{t}$ events (from MC predictions) and is used for studying the modelling of $t\bar{t}$ events. For the measurement of the cross-section with additional b-jets, the events are defined as the subset of the baseline selection where at least three or four jets must be b -tagged. A MC based cut-flow is presented in Table 5.3 to show the effect of each event selection requirement leading to 95% pure $t\bar{t}$ events.

Object definitions for detector-level event selection	
Lepton selection	
Leptons	$e\mu$ pair (exactly 1 e and 1 μ with opposite charge)
Transverse momentum ${p_{\mathrm{T}}}^{e,\mu}$	$> 27\,\mathrm{GeV}$
Pseudorapidity $\lvert\eta_{e,\mu}\rvert$	< 2.47 but not $1.37 < \lvert\eta\rvert < 1.52$
Jet Selection	
Identification of b-jets	b-tagging by MV2c10 tagger at 77%
Number of b-tagged jets, $N_{b\text{-jets}}$	≥ 2
Transverse momentum $p_{\mathrm{T}}^{b_1}$	$> 25\,\mathrm{GeV}$
Pseudorapidity $\lvert\eta_b\rvert$	< 2.5
Additional requirement	
No jet-electron or jet-muon pair with $\Delta R < 0.4$	

Table 5.2.: Summary of the detector level event selection criteria used in this analysis for the baseline selection.

Processes		Event Selection Requirements				
		Pre-selection	Dilepton	$e\mu$	Opposite-sign	$\geq 2b$-jets
Signal	$t\bar{t}$	1.38	1.26	31.11	31.28	96.37
	$t\bar{t}V$	0.04	0.04	0.97	0.98	0.26
	$t\bar{t}H$	0.001	0.001	0.02	0.02	0.08
Background	Wt	1.11	0.31	7.64	7.68	3.21
	Drell-Yan	96.81	98.02	56.96	57.013	0.07
	Diboson	0.66	0.36	3.29	3.03	0.02

Table 5.3.: Percentage of events for signal and background after each event selection requirement as computed purely from MC predictions.

5.3. Background Estimation

It can be seen from Table 5.3 that after the baseline selection of an $e\mu$ pair and at least 2 b-tagged jets, only a small fraction of background events from non-$t\bar{t}$ production processes remain in the sample. The signal purity is > 95% with the dominant Wt background corresponding to $\sim 3.2\%$. Depending on the type of background, different approaches such as data-driven approach or MC estimation are used to determine their contributions. A "data-driven" approach refers to the background estimation method which utilises both the data and the MC simulation. Some background processes, like non-prompt and fake leptons, are almost impossible to simulate using MC generators because they involve the prediction of instrumental effects of the detector which is a difficult task. A data-driven technique is employed in such cases. The processes involving the tails of distributions which are not very precisely modelled, such as Drell-Yan, also require data-driven approach.

The backgrounds coming from processes such as single top and diboson productions are well described by the simulations and are estimated directly from the MC samples. The expected number of background events from a process is calculated by:

$$N_{\text{expected}} = L \times \sigma \times \frac{N_{\text{selected}}}{N_{\text{total}}} \tag{5.1}$$

where L is the integrated luminosity used in the analysis, which is 36.1 fb^{-1}. σ is the cross section and $\frac{N_{\text{selected}}}{N_{\text{total}}}$ is the ratio of the number of events passing the selection criteria to the total number of events of the background process under consideration.

5.3.1. Non-prompt and Fake Leptons

The non-prompt (NP) and fake lepton background together is referred to as fake background in this text for simplicity. Majority of the background with at least one fake lepton in the selected sample arises from $t\bar{t}$ semi-leptonic decay and $W+$ jets events. In such cases one real lepton comes from the W boson and a hadronic jet is misidentified as the second lepton. The other sources which contribute to this background are the non-prompt leptons originating from photon conversions, heavy flavour hadron decays or in-flight decays of mesons. This background is estimated from a data-driven approach called *fake-factor* method as employed in Reference [165].

In the fake factor method, the samples with a same-sign (SS) $e\mu$ pair and two or at least three b-tagged jets are used to define the *control region*. A control region is selected such that it is enriched with that particular background process and is orthogonal to the signal region. Being orthogonal to the signal region implies that no signal is expected in this region and it can be used to understand the background. All the other selections are the same as mentioned in the event selection. The SS samples are dominated by events with one real and one fake lepton as can be seen in Table 5.4. It also shows that the number of fake events in the opposite-sign (OS) and same-sign (SS) samples are comparable, at least around a factor of two, and that the simulation predicts the number of same-sign events within about 10% as observed in data for the baseline selection. Therefore, the fake lepton contribution to the OS samples are estimated using the SS events in data and simulation.

Composition of the Same-Sign Sample

The composition of OS and SS events in the simulated sample is summarised in table 5.4. The $t\bar{t}$ sample here consists of both the di-leptonic and semi-leptonic decay modes in order to account for all fake lepton contributions. The fake leptons are categorized depending on the origin of the leptons. "Conversion leptons" refer to the leptons originating from converted photons or π, K meson decays, "heavy flavour" implies to the leptons coming from the heavy flavour hadron decay and "other" consists of all remaining un-identified sources.

The contribution of fake leptons from different physics processes are shown in Table 5.5. The SS events are dominated by the contributions from $t\bar{t}$, $t\bar{t}H$ and $t\bar{t}V$ processes. Only about 4% (10%) contributions come from other processes e.g. diboson and single top for $2b$ ($\geq 3b$) selection. According to the simulation, the fake lepton background stems dominantly from mis-identified electrons and much less from mis-identified muons. Therefore, it can be inferred that the highest contribution of fake leptons comes from the di-leptonic $t\bar{t}$ events ($t\bar{t} \rightarrow e\mu\nu\nu b\bar{b}$) where

	2 b-tagged jets			≥ 3 b-tagged jets		
Components	OS	SS	f (OS/SS)	OS	SS	f (OS/SS)
Conversion e	438 $\pm$ 21	204 $\pm$ 14		16 $\pm$ 4	9 $\pm$ 3	
Heavy flavor e	12 $\pm$ 3	10 $\pm$ 3		0.7 $\pm$ 0.8	0.8 $\pm$ 0.8	
Other e	25 $\pm$ 5	1.4 $\pm$ 1.2		1.0 $\pm$ 1.0	1.3 $\pm$ 1.1	
Conversion μ	3.6 $\pm$ 1.9	0.5 $\pm$ 0.7		0.1 $\pm$ 0.2	0.0 $\pm$ 0.1	
Heavy flavor μ	21 $\pm$ 5	9.5 $\pm$ 3.1		0.7 $\pm$ 0.8	0.3 $\pm$ 0.6	
Other μ	18 $\pm$ 4	1.0 $\pm$ 1.0		0.7 $\pm$ 0.8	0.1 $\pm$ 0.4	
Fake leptons	519 $\pm$ 23	227 $\pm$ 15	2.3 $\pm$ 0.2	19 $\pm$ 4	10 $\pm$ 3	1.6 $\pm$ 0.7
Prompt	75228 $\pm$ 274	199 $\pm$ 14		3209 $\pm$ 57	13 $\pm$ 4	
Total MC	75747 $\pm$ 275	426 $\pm$ 21		3228 $\pm$ 57	25 $\pm$ 5	
Total Data	76425	468		3809	38	

Table 5.4.: The expected number of events with fake leptons in 2 and ≥ 3 b-tagged jets in opposite- and same-sign (OS and SS) $e\mu$ samples, broken down into different categories based on their origins as described in the text. The fake factor, f, is the ratio of the number of OS and SS events for the total fake leptons. The uncertainties are due to simulation statistics.

a photon is produced by bremsstrahlung from the electron from the decay chain $t \to W \to e$ which further gets converted to an electron-positron pair i.e. ($e^- \to e^-\gamma \to e^-e^+e^-$).

The second-largest contribution comes from heavy flavour decays, giving rise to both fake electron and fake muon contributions. These backgrounds are dominated by semi-leptonic $t\bar{t}$ events, with one real lepton and a second lepton from the decay of a bottom or charm hadron, which happens to pass the lepton isolation cuts. There is also a small contribution from W+jets events with heavy flavour. These backgrounds are heavily suppressed in the two b-tagged sample, where both the b-quarks produced in the top decays have already produced well-defined b-tagged jets.

The small contribution from other fake electrons includes contributions from hadrons misidentified as electrons in semi-leptonic $t\bar{t} \to \mu\nu q\bar{q}b\bar{b}$ and $W \to \mu\nu$+jets events, consistent with being charge symmetric. A larger opposite-sign contribution from other fake muons comes from $Z \to \mu\mu$+jets and $t\bar{t} \to \mu\mu\nu\nu b\bar{b}$ where an isolated muon is misidentified as an electron.

Figure 5.3 shows the kinematic distributions of the electrons and muons in the SS events with the fraction of contributions of fake leptons in the sample.

Process	$2b$	$\geq 3b$
$t\bar{t}$	116 ± 11	3.9 ± 2.0
$t\bar{t}V$	71 ± 8	7.9 ± 2.8
$t\bar{t}H$	3.3 ± 1.8	0.4 ± 0.7
Wt	4.2 ± 2.1	0.7 ± 0.8
Dibosons	3.8 ± 2.0	0.5 ± 0.7
$Z/\gamma*$+jets	0.02 ± 0.14	0.0 ± 0.0
Total	199 ± 14	13 ± 4

Table 5.5.: Contribution from different sources for 2 b-tagged and at least 3 b-tagged prompt SS events in the $e\mu$ channel. Uncertainties are statistical only.

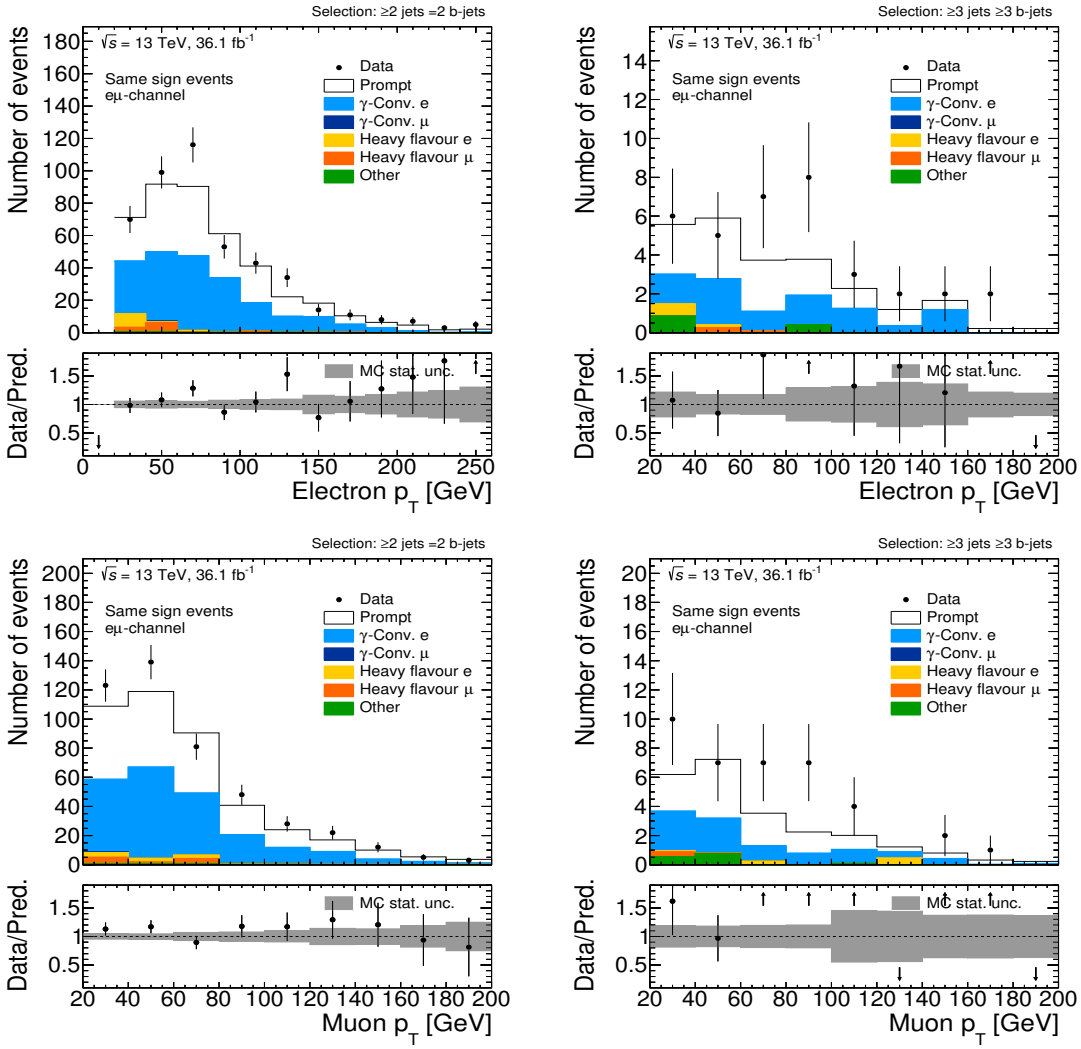


Figure 5.3.: Electron p_T spectra of events with a $e\mu$ pair with the same sign in data compared to simulations. Events with exactly two b-tagged jets (left) and events with at least three b-tagged jets (right).

Fake Lepton Background Estimate and Uncertainties

The contributions of events with misidentified leptons are estimated in events with exactly two or at least three b-tagged jets. The separation in number of b-tagged jets is done to cover any potential difference related to leptons produced in b-hadron decays. As mentioned, the contributions of events with fake leptons are estimated using the same-sign event counts. The number of fake events is determined by subtracting the estimated prompt SS contributions from the SS events in data and then multiplied by the fake factor (Equation 5.2). The fake factor, f (Equation 5.3), is the ratio of opposite- to same-sign fake-leptons predicted from the MC simulations.

$$N_{\text{fake}} = (N^{\text{SS}}_{\text{data}} - N^{\text{SS}}_{\text{prompt,MC}}) \cdot f \tag{5.2}$$

$$f = \frac{N^{\text{OS}}_{\text{fake,MC}}}{N^{\text{SS}}_{\text{fake,MC}}} \tag{5.3}$$

N_{fake} refers to the number of fake background events, $N^{\text{SS}}_{\text{data}}$ refers to the number of SS lepton events in data and $N^{\text{SS}}_{\text{prompt,MC}}$ is the number of events predicted from MC simulation where one of prompt leptons is reconstructed with the wrong sign. $N^{\text{OS}}_{\text{fake,MC}}$ and $N^{\text{SS}}_{\text{fake,MC}}$ are the number of mis-identified events estimated from MC simulation with OS and SS leptons, respectively. $N^{\text{SS}}_{\text{data}}$ and $N^{\text{SS}}_{\text{prompt,MC}}$ are evaluated for each bin for each of the observables, while f is determined for different b-tag jet multiplicity. The value of f is $2.3 \pm 0.2\,(\text{stat})$ for the $2\,b$-tagged jet bin and $1.7 \pm 0.7\,(\text{stat})$ for $\geq 3\,b$-tagged jet bins. The values are compatible within uncertainties, no strong dependence on the b-tagged jet multiplicity is observed. Therefore the fakes estimated in events with three b-tagged jets can also be applied to the events with $4\,b$-tagged jets where the low statistic does not allow for an estimation of fakes.

Looking at the number of predicted OS and SS events in Table 5.4, it can be inferred that the OS/SS ratio for each contribution is different and would lead to very high values in some cases along with large uncertainties due to limited simulation statistics. In addition, the predictions for $N^{\text{SS}}_{\text{prompt,MC}}$ are sensitive to the cross-section of $t\bar{t}$ production, $t\bar{t}$ with W or Z bosons and diboson production with additional b-tagged jets which come with the underlying uncertainties in the measurements such as lepton, jet and b-tagging uncertainties. Table 5.4 also shows that the SS events in data differ from MC by a factor 1.08 (1.5) for the $2b$ ($\geq 3b$) bin. This apparent mis-modelling of the fakes source is taken into account by applying an uncertainty of 100% on the fakes estimate.

5.3.2. Drell-Yan

A very small fraction of background comes from the Drell-Yan (DY) process, where a Z boson or virtual photon decays to two same flavour leptons with opposite charge. The events originating from $Z/\gamma^* \to \tau\tau \to e\mu$ with two and three or more b-tagged jets contaminate the selected events. This background is estimated from the SHERPA2.1 simulated samples (Section 3.4.4) with a data-driven normalisation as described in the following.

The normalisation is derived by comparing MC predictions to data in the same flavor dilepton samples $(ee, \mu\mu)$ in the Z control region after subtraction of the background as shown in Equation 5.4. The control region (CR) is defined around the Z -mass peak ($81 < m_{ll} < 101$ GeV), where m_{ll} is the di-lepton invariant mass. The lepton selection follows the object definition as in the main analysis, mentioned in Section 5.2. The normalization or the Drell-Yan scale factor, $\mathrm{SF_{DY}}$, is therefore calculated as

$$\mathrm{SF_{DY}} = \frac{N^{\mathrm{CR}}_{\mathrm{data}} - N^{\mathrm{CR}}_{\mathrm{non-DY,MC}}}{N^{\mathrm{CR}}_{\mathrm{DY,MC}}} \tag{5.4}$$

where $N^{\mathrm{CR}}_{\mathrm{data}}$ is the number of DY events in data, $N^{\mathrm{CR}}_{\mathrm{DY,MC}}$ and $N^{\mathrm{CR}}_{\mathrm{non-DY,MC}}$ are MC predicted number of DY and non-DY events in the control region respectively. The scale factors are derived in the ee and $\mu\mu$ channels separately for 2 b-tagged jets events as shown in Figures 5.4-5.5.

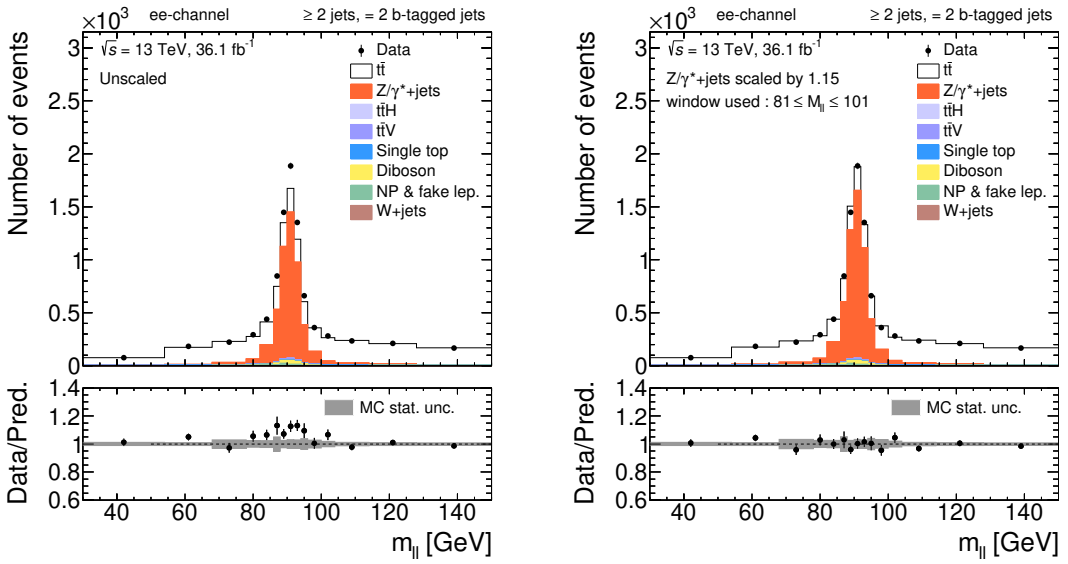


Figure 5.4.: Dilepton invariant mass for the ee channel in events with two b-tagged jets without (left) and with (right) the Drell-Yan scale factors applied on $t\bar{t}$ events.

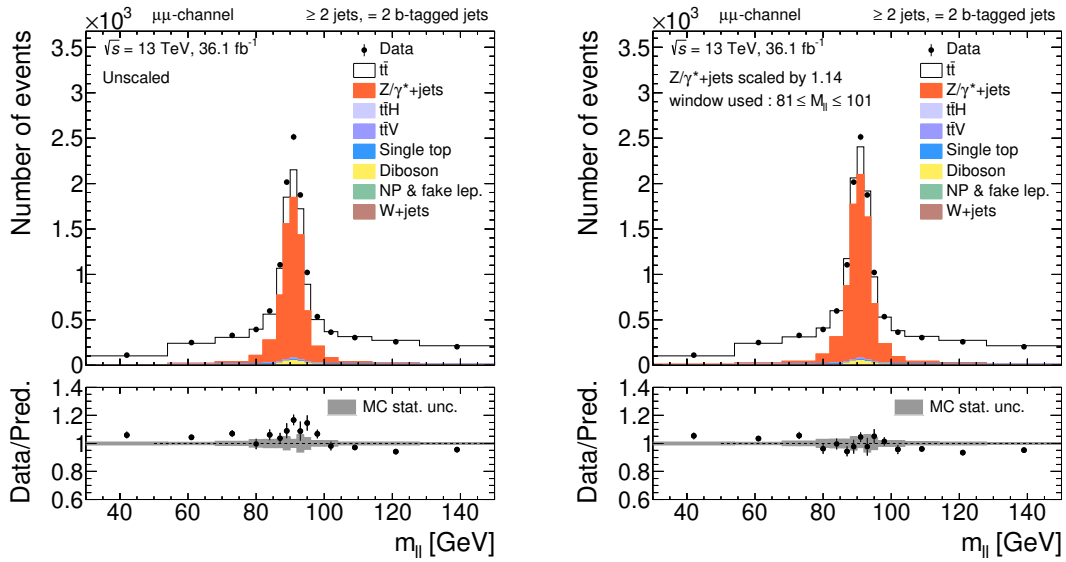


Figure 5.5.: Dilepton invariant mass for the $\mu\mu$ channel in events with two b-tagged jets without (left) and with (right) the Drell-Yan scale factors applied on $t\bar{t}$ events.

The SF_{DY} for the $e\mu$ channel is derived by combining the events from ee and $\mu\mu$ channels and the distributions are shown in Figure 5.6. Since it is not expected that the production of multiple additional b-jets in Z+jets events is described precisely, the analysis is performed separately for events with 2 b-tagged jets and three or more b-tagged jets. The SF_{DY} for events ≥ 3 b-tagged events are derived using the same procedure and the dilepton invariant mass is presented in Figure 5.7. The scale factors are summarised in Table 5.6 along with the quoted uncertainties.

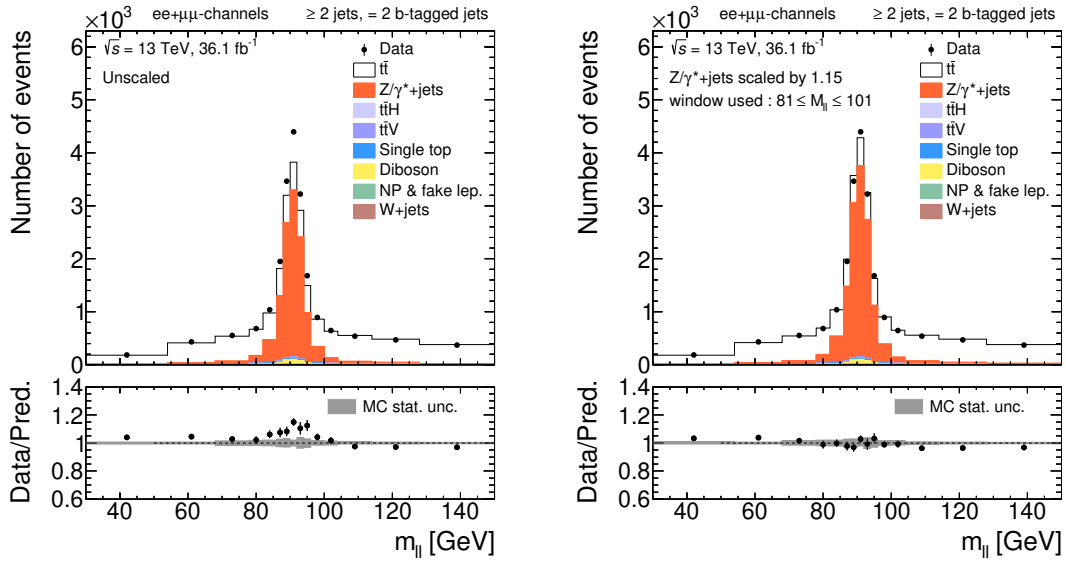


Figure 5.6.: Dilepton invariant mass for the combined ee and $\mu\mu$ channels in events with two b-tagged jets without (left) and with (right) the Drell-Yan scale factors applied on $t\bar{t}$ events.

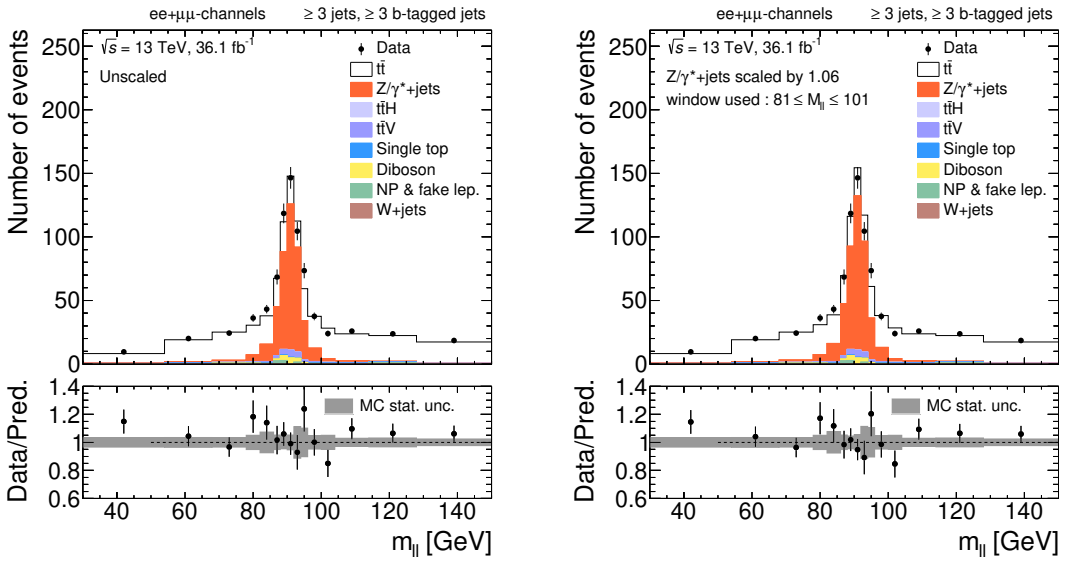


Figure 5.7.: Dilepton invariant mass for the combined ee and $\mu\mu$ channels in events with at least three b-tagged jets without (left) and with (right) the Drell-Yan scale factors applied on $t\bar{t}$ events.

Channel	2b-tagged-jets	$\geq$ 3 b-tagged-jets
ee	1.15 $\pm$ 0.02 $\pm$ 0.04	0.81 $\pm$ 0.10 $\pm$ 0.06
$\mu\mu$	1.14 $\pm$ 0.03 $\pm$ 0.04	1.24 $\pm$ 0.09 $\pm$ 0.05
$e\mu$	1.14 $\pm$ 0.02 $\pm$ 0.04	1.05 $\pm$ 0.07 $\pm$ 0.25

Table 5.6.: The Drell-Yan scale factor computed from the data-driven method in ee and $\mu\mu$ channels for 2 and at least 3b-tagged jets. For the $e\mu$ channel the SF_{DY} is computed by combining the events from the other two channels. The second and third terms are the statistical and systematic errors respectively.

As can be seen in Figures 5.6 and 5.7, the effect of the scale factors on the 2 b-tagged jet selection is minimal, while it improves significantly the description of the tails in the 3 b-tagged jet selection leading to a good description, as expected. The scale factors obtained from ee and $\mu\mu$ combined sample is applied to the simulated DY events to get the resulting contribution from the DY background.

The systematic uncertainties on these scale factors are derived from varying the cuts defining the control region. These systematics sources are attributed to detector related background that can affect the mass resolution and to the background physics processes that contribute to di-lepton events in the Z mass window. The mass window is varied by $\pm$ 6 GeV and the amount of background predicted by MC simulation is scaled by 25%. For the $e\mu$ channel an additional systematics uncertainty is associated due to difference in scale factors obtained from ee and $\mu\mu$ channels. It can be noted that for the 2b selection this uncertainty is negligible

since both *ee* and $\mu\mu$ SF_{DY} are quite similar (Table 5.6). However, for the $\geq 3b$ selection it is the biggest source of uncertainty due to the large difference observed in the *ee* and $\mu\mu$ SF_{DY}.

5.3.3. Single Top

The production of single top in association with a real W boson (Wt-channel) is the dominant background in this analysis contributing only 3.16%. The NLO real emission contribution to Wt has the same configuration as the LO $t\bar{t}$ production which causes interference between them. There are two approaches to deal with this interference, namely *diagram removal* (DR) and *diagram subtraction* (DS) [166]. The DR method removes resonant effects of Wt from $t\bar{t}$ at the amplitude level and the DS at the cross-section level. The Wt single top background is determined from the MC samples described in Section 3.4.4 and normalised to the approximate NNLO cross-section of 71.7 ± 3.8 pb as measured in Reference [167].

The sensitivity of the MC predictions to this interference is studied by comparing the predictions of POWHEG with the DR and DS schemes [166]. A relative difference of 30% is found and considered as systematic uncertainty on the normalisation. Due to the smallness of this background, especially in the higher jet multiplicity bins an additional uncertainty on the shape is considered negligible.

5.3.4. Diboson

The diboson production is the smallest background for the signal process. This background is also estimated using the corresponding MC predictions (Section 3.4.4) and normalised to the approximate NLO cross-section. The shape and normalisation are estimated using SHERPA2.1 with an uncertainty of 50% on the cross-section following the procedure of [168]. This uncertainty covers possible variations of the normalisation and the shapes due to the choices of the PDF, the scales (renormalisation, factorisation, matching scales) and the electroweak parameters.

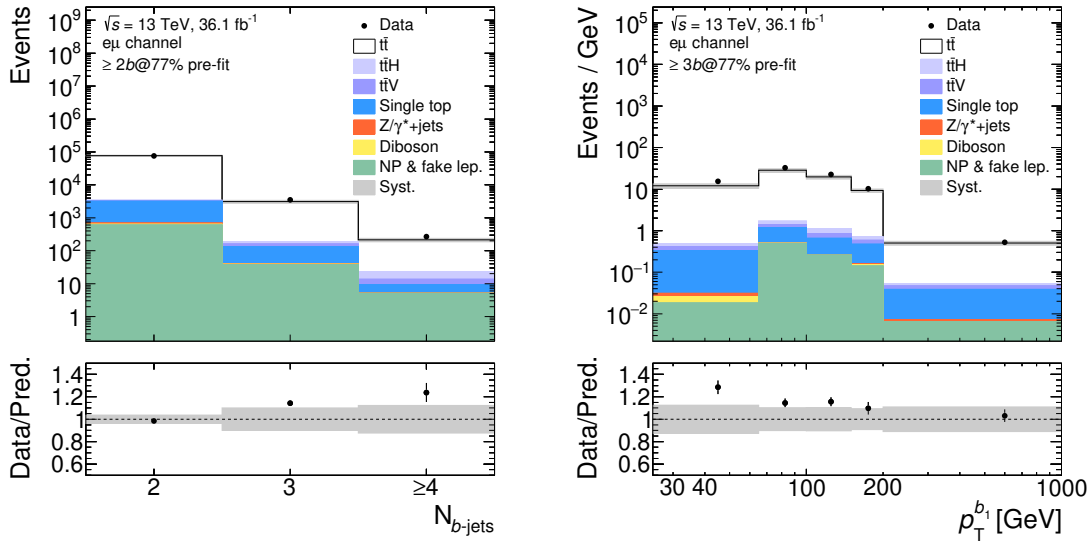
5.4. Event Yield

The resulting event yield after the standard event selection and background estimates are shown in Table 5.7 for different b-jet multiplicities. The number of events fulfilling the baseline selection is well described by the prediction as seen in Table 5.7 and Figure 5.8a. However, the number of events with more than two b-tagged jets is slightly underestimated as shown in

Figure 5.8. The non-$t\bar{t}$ background is very low in all the three selections amounting to around 4% in each case.

Process	$2b$-jet	$\geq 3\,b$-jets	$\geq 4\,b$-jets
$t\bar{t}$	74214 ± 272 ± 2872	3107 ± 56 ± 308	193 ± 14 ± 26
$t\bar{t}V$	190 ± 14 ± 38	33.5 ± 5.8 ± 6.7	4.4 ± 2.1 ± 0.9
$t\bar{t}H$	45.3 ± 6.7 ± 22.7	36.5 ± 6.0 ± 18.2	9.4 ± 3.1 ± 4.7
Signal ($t\bar{t}$ + $t\bar{t}V$ + $t\bar{t}H$)	74449 ± 273 ± 2873	3177 ± 56 ± 308	207 ± 14 ± 26
Wt	2455 ± 50 ± 537	96 ± 10 ± 30	4.1 ± 2.0 ± 1.4
Fake leptons	604 ± 25 ± 604	43.0 ± 6.6 ± 43.0	5.1 ± 2.3 ± 5.1
$Z/\gamma*$+jets	52.8 ± 7.3 ± 13.2	1.30 ± 1.14 ± 0.32	0.07 ± 0.27 ± 0.02
Dibosons	37.6 ± 6.1 ± 18.8	1.02 ± 1.01 ± 0.51	0.00 ± 0.00 ± 0.00
Total background	3149 ± 56 ± 808	141 ± 12 ± 53	9.2 ± 3.0 ± 5.3
Expected	77599 ± 279 ± 2984	3318 ± 58 ± 313	216 ± 15 ± 27
Observed	76425	3809	267

Table 5.7.: Event yields for different b-tagged jet multiplicity for data and MC fulfilling the baseline selection. The quoted errors are the statistical and systematic uncertainties in the predictions due to experimental sources.



(a) **(b)**

Figure 5.8.: Data to MC comparison of (a) number of b-tagged jets for events fulfilling the $2b$ baseline selection and (b) p_{T} of the leading b-jet for events fulfilling the $\geq$ $3b$ selection. The systematic uncertainty band (grey) includes all uncertainties from experimental sources.

5.5. Correction for Flavour Composition of Additional Jets

The measurement of $t\bar{t}+b$-jets production consists of background from other $t\bar{t}$ processes where at least one c-jet or light flavour-jet is misidentified as a b-jet by the b-tagging algorithm [155]. This is referred to as $t\bar{t}c$ and $t\bar{t}l$ background.

A simulation based study using the nominal $t\bar{t}$ sample is performed to illustrate the fraction of this background. The reconstructed jets which are tagged as b-jets using the MV2c10 b-tagging algorithm (Section 4.5) are referred to as reconstructed b-jets or b-tagged jets. However, the truth or particle-level jets from the MC generated sample are called truth b-jets. As mentioned in Section 4.8, the flavour of the truth jet is determined by ghost matching of b- and c-hadrons [164]. To understand this background, events with two and three reconstructed jets which are tagged as b-jets are taken. Each of the b-tagged jet is then matched to a truth particle jet with minimum ΔR of 0.4 between the reconstructed and truth jet. It should be noted that since the b-tagging procedure cannot be 100% accurate, some of the b-tagged jets can originate from a c- or light-flavour quarks and are matched with truth level c- or light-jets. Figure 5.9 shows the truth event composition in the events selected with 2 b-tagged jet and 3 b-tagged jets. Figure 5.9a shows that the events fulfilling the 2b selection are 97% pure i.e. both b-tagged jets originate from a jet containing a b-hadron. Whereas, only about 50% of the events with 3 b-tagged jets have three b-jets on generated particle level (Figure 5.9b). About 28% of events contain 2 generated b-jets and one c-jet and 20% of events contain 2 generated b-jets and one light jet.

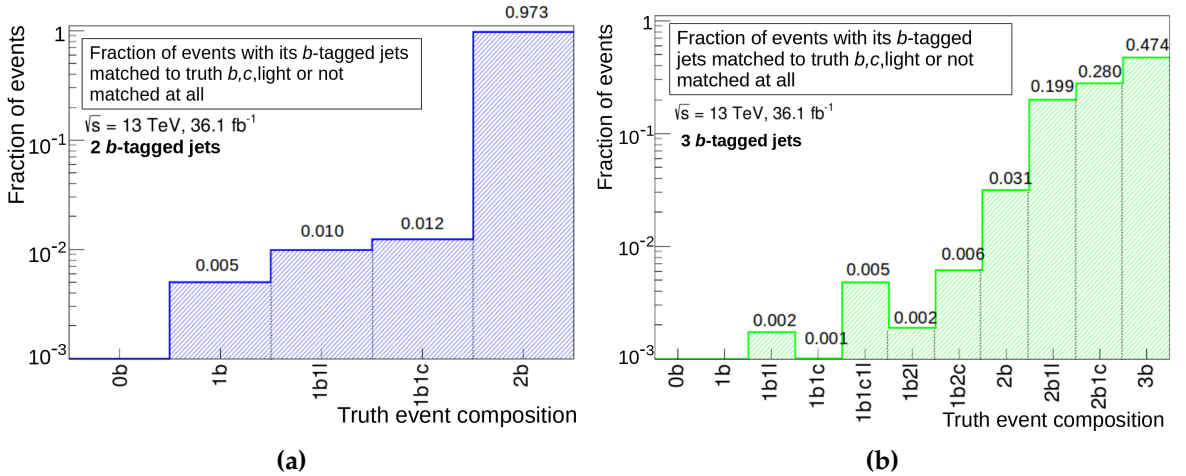


Figure 5.9.: Composition of $t\bar{t}+b$-jets events in categories of different flavours of jets generated in POWHEG+PYTHIA 8 sample. The events fulfil the baseline selection and contain different numbers of b-tagged jets. (a) Distributions with exactly 2 b-tagged jets and (b) exactly 3 b-tagged jets.

The measurement of $t\bar{t}+b$-jets production is therefore tightly bound to the determination of the background from $t\bar{t}c$ and $t\bar{t}l$ processes. The cross-section measurement of $t\bar{t}$ production with additional jets has been performed with 10% (16%) uncertainty for events with two (three) additional jets [169]. However, the flavour of the additional jets are not determined in these measurements. Due to the lack of precision measurements of these processes, a template fit to data is employed to extract the $t\bar{t}b$ signal yields and estimate the background from $t\bar{t}c$ and $t\bar{t}l$ as described in the following.

Events having at least 3 b-tagged jets at 77% b-tagging working point (WP)[2] are selected for the $t\bar{t}$, $t\bar{t}H$ and $t\bar{t}V$ MC samples as the signal includes all three contributions. The information of particle-level jet flavor is then extracted to identify the event as signal (i.e. containing at least 3 b-jets at particle level) or background (one or more mistagged light- or c-jets). The particle-level event categorization is done purely based on the generated particle-level jets without requiring matching to the reconstructed-level jets. The event categories are defined as

- $t\bar{t}b$ = Number of events with ≥ 3 b-tagged reconstructed jets which have three or more particle level b-jets (≥ 3 b-jets).
- $t\bar{t}c$ = Number of events with ≥ 3 b-tagged reconstructed jets, which have less than three b-jets and at least one c-jet at particle level (< 3 b-jets and ≥ 1 c-jets)
- $t\bar{t}l$ = Number of events with ≥ 3 b-tagged reconstructed jets, which have less than three b-jets and no c-jet at particle level (events not falling in $t\bar{t}b$ and $t\bar{t}c$ criteria).

Thereby, the MC predicted $t\bar{t}$ events with ≥ 3 b-tagged jets are divided into two templates of signal ($t\bar{t}b$) and background ($t\bar{t}c$ +$t\bar{t}l$). Now, the reconstructed-level jets are ordered by their b-tagging discriminant value for the data, $t\bar{t}b$, $t\bar{t}c$ +$t\bar{t}l$ and non-$t\bar{t}$ background samples. The jet with the 3$^{\text{rd}}$ highest discriminant value is taken and its discriminant value is divided into three bins corresponding to the efficiency between 77% and 0%. Three templates are created from the $t\bar{t}$, $t\bar{t}H$, $t\bar{t}V$ MC samples to account for $t\bar{t}b$ and $t\bar{t}c$ +$t\bar{t}l$ events and one template is created from the sum of all backgrounds described in Section 5.3. Since the templates contain lower statistics and number of events are to be divided into discrete number of bins, the templates are fitted to data using a binned maximum-likelihood fit with a Poisson distribution. The likelihood based on Poisson statistics is given by

$$L(\vec{\alpha}|x_1,....,x_n) = \prod_k^n \frac{\nu_k(\vec{\alpha})^{x_k} e^{-\nu_k(\vec{\alpha})}}{x_k!} \tag{5.5}$$

where $\vec{\alpha}$ is the number of free parameters to be fitted, x_k is the number of events in bin k of the data template and $\nu_k(\vec{\alpha})$ is the expected number of events which depends on the number

[2]Please refer to Section 4.5 for the b-tagging details and related specific terms used frequently in the following text

of free parameters. There are two free parameters used in this flavour fit: α_b and α_{cl}. The expected number of events in bin k is determined by

$$\nu_k(\alpha_b, \alpha_{cl}) = \alpha_b N^k_{t\bar{t}b} + \alpha_{cl}(N^k_{t\bar{t}c} + N^k_{t\bar{t}l}) + N^k_{\text{non}-t\bar{t}} \tag{5.6}$$

where $N^k_{t\bar{t}b}$, $N^k_{t\bar{t}c}$, $N^k_{t\bar{t}l}$ and $N^k_{\text{non}-t\bar{t}}$ represents the number of events in bin k of the $t\bar{t}b$, $t\bar{t}c$, $t\bar{t}l$ and non-$t\bar{t}$ background templates. $t\bar{t}c$ and $t\bar{t}l$ are taken together to get a stable fit in the low statistics data sample. The scale factors obtained from the flavour fit are $\alpha_b = 1.37 \pm 0.06$(stat) and $\alpha_{cl} = 1.05 \pm 0.04$(stat). After applying the correction factors obtained by the flavour fit, a good agreement is reached over the full spectrum as shown in Figure 5.10. The bins 3, 4 and 5 correspond to the b-tagging efficiency of 77%-70%, 70%-60% and <60%, respectively. The dotted line shows the template before scaling and the blue shaded region shows the fitted template.

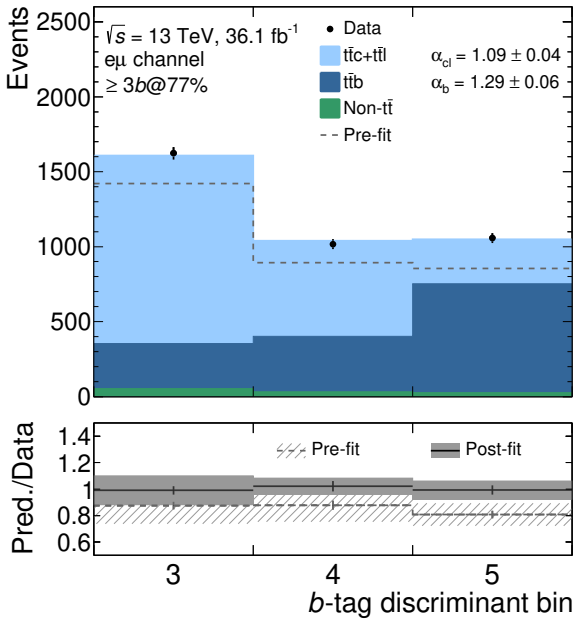


Figure 5.10.: MV2c10 discriminant output for the 3rd b-tagged jet in the event selected with at least three b-tagged jets at 77% WP. The b-tagged jets are ordered according to their MV2c10 discriminant value. The data are shown in comparison to MC separated into the different event categories.

The choice of 77% working point in the analysis is used to enrich the sample with b-jets. Going to higher efficiency b-tagging working points decreases the c-jet and light-jet rejection rates which contaminates the sample with non-$t\bar{t}b$ events. The fits are performed with the same working point to avoid extrapolation of background shapes determined outside the selected region to the signal region. Although to estimate the stability of the template fit a variation of the fitting range and a variation of the fractions of c-jets and light-jets in the combined template is done. Since the $t\bar{t}c$ and $t\bar{t}l$ backgrounds are determined in a single fit, the uncertainty in this result is determined by changing the sample composition.

Figure 5.11 shows a fit using 4 and 5 bins in the b-tagging discriminant. In this case the 2 b-tagged jets are b-tagged at 77% WP while looser cuts are allowed to the 3rd b-tagged jets. The 4 bin fit includes 3rd jet with a b-tagging efficiency between 77%-85% and adding the 5^{th} bin implies 100% efficiency which means that the jets are not required to be b-tagged at all. Both show a maximal deviation of about the statistical error on α_b. With these looser selections the values α_c varies by about 40% and this is used as a systematic uncertainty in the $t\bar{t}c$ template. However, the values of α_l remain consistent with the statistical uncertainty in fit with looser selections. Therefore, a fit is performed by varying the input c-templates by $\pm 40\%$ before performing the fit (Figure 5.12). As normalisation is preserved for $t\bar{t}c$ +$t\bar{t}l$ template, therefore when $t\bar{t}c$ is scaled up by 40%, the $t\bar{t}l$ is scaled down by 30%. After propagating the uncertainty in $t\bar{t}c$ template, the value of α_b is found to change by $\pm 11\%$ while the value for α_{cl} changes by $\pm 7\%$.

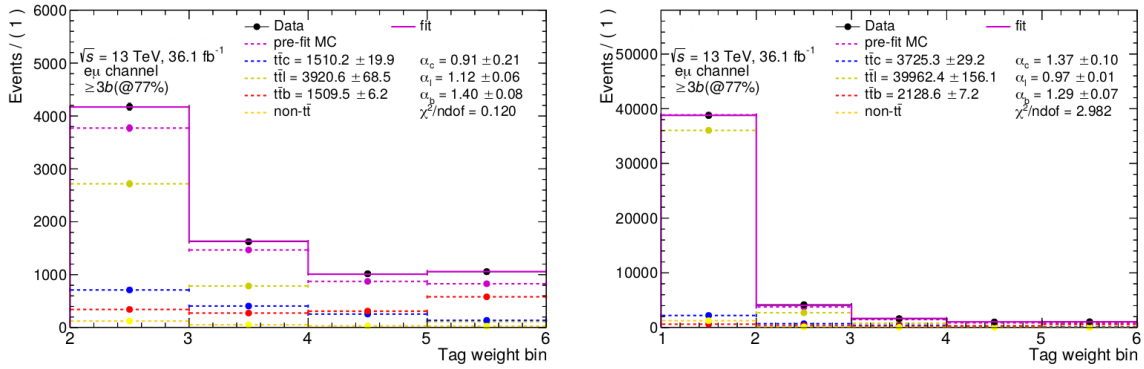


Figure 5.11.: MV2c10 discriminant output for the 3rd jet ordered by MV2c10 discriminant value in the events with at least 2 b-tagged jets at 77% working point. The fit range is considered to include looser efficiencies for the 3rd jet: (left) fit range covers 85%-0% WP (4 bins) and (right) fit range covers 100%-0% WP (5 bins). The fits are performed after considering separate templates for $t\bar{t}c$ and $t\bar{t}l$.

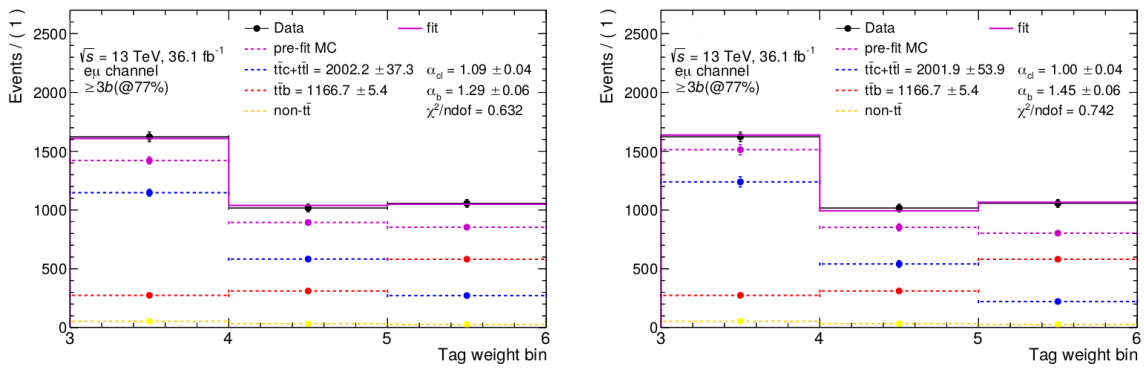
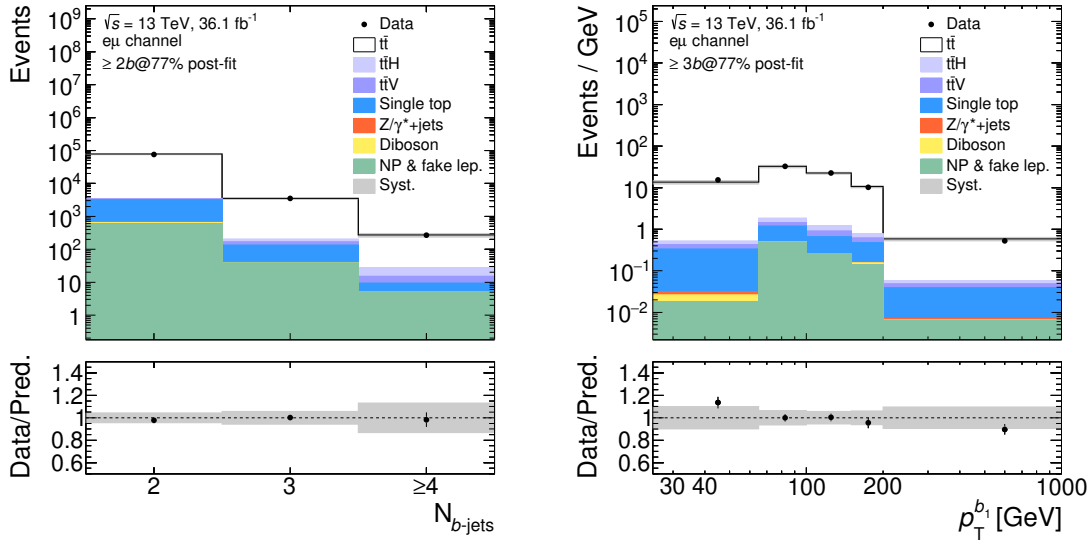


Figure 5.12.: MV2c10 discriminant output for the 3rd b-tagged jet at 77% working point in the events with at least 3 b-tagged jets at 77% WP and the fit is performed after the c-contribution is changed in the $t\bar{t}c$+$t\bar{t}l$ template by 40% upward (left) and 40% downward (right).

Figure 5.13 shows that good data-MC agreement is reached after the flavour fit for the b-jet multiplicity as well as the p_{T} of the leading b-jet as compared to the pre-fit plots in Figure 5.8.



(a) (b)

Figure 5.13.: Data to MC comparison of (a) number of b-tagged jets for events fulfilling the $2b$ baseline selection and (b) p_{T} spectrum of the leading b-jet for events fulfilling the $\geq 3b$ selection after applying the scale factors to the $t\bar{t}$ MC obtained by the flavour fit. The systematic uncertainty band (grey) includes all uncertainties from experimental sources.

5.6. Observables

This section outlines the observables which are chosen to present the differential cross-section as these variables are discussed very frequently in the process of getting to the results. To investigate the properties of the $t\bar{t}$ system with additional b-jets, it is very reasonable to analyse the distribution of the number of b-jets referred to as b-jet multiplicity. The second set of variables of interest is the distribution of the p_{T} of the b-jets. The b-jets are arranged in the order of descending b-jet p_{T} and the p_{T} spectra of the first three leading b-jets are presented for the $t\bar{t}b$ process.

In addition, kinematic distributions of global event properties are also measured such as the scalar sum of the p_{T} of all objects mentioned in the event selection: electron, muon and b-jets and the scalar sum of p_{T} of all the jets.

In order to investigate the properties of the $b\bar{b}$ system, two approaches are used: one is the selection of the b-jet pair consisting of the two leading p_{T} b-jets and the other system is

built based on the two closely lying b-jets. The system of two leading b-jets is denoted by b_1b_2 system. The distance between the two b-jets is determined based on the ΔR between them. The system having minimum ΔR between them is denoted by $bb_{\Delta\text{min}}$ system. The observables studied in both the cases are invariant mass, p_{T} and angular distance between them.

To summarize, the variables used in this analysis to assess the results are:

- b-jet multiplicity ($N_{b\text{-jets}}$)
- p_{T} distribution of the highest p_{T} b-jet called leading b-jet ($p_{\text{T}}^{b_1}$). Similarly, the p_{T} spectra of second and third leading b-jets ($p_{\text{T}}^{b_2}$ and $p_{\text{T}}^{b_3}$)
- Scalar sum of the p_{T} of all objects in the event is defined as H_{T} and of all jets in the event as $H_{\text{T}}^{\text{had}}$
- Angular distance between the two highest p_{T} b-jets ($\Delta R_{b_1b_2}$), invariant mass and p_{T} of two leading b-jets ($m_{b_1b_2}$ and p_{T,b_1b_2})
- Invariant mass, transverse momentum and angular correlation between two closely lying b-jets ($m_{bb}^{\Delta\text{min}}$, $p_{T,bb}^{\Delta\text{min}}$ and $\Delta R_{bb}^{\Delta\text{min}}$)

5.7. Data and MC Comparison

The distributions of the b-jet multiplicity and transverse momenta of leading b-jet are already shown in the text above; both before and after applying the flavour fit scale factors. Figures 5.14-5.16 present the data and MC comparison of other observables at the detector-level. It can be observed that the event selection discussed in Section 5.2 yields a sample dominated by $t\bar{t}$ events. The distributions on the left show that the MC samples are not well modelled for some observables. The agreement between the data and MC improves after the scale factors obtained from the flavour fit are applied (plots shown in right side). The improved agreement is shown to justify that the distributions can be unfolded after applying these correction factors.

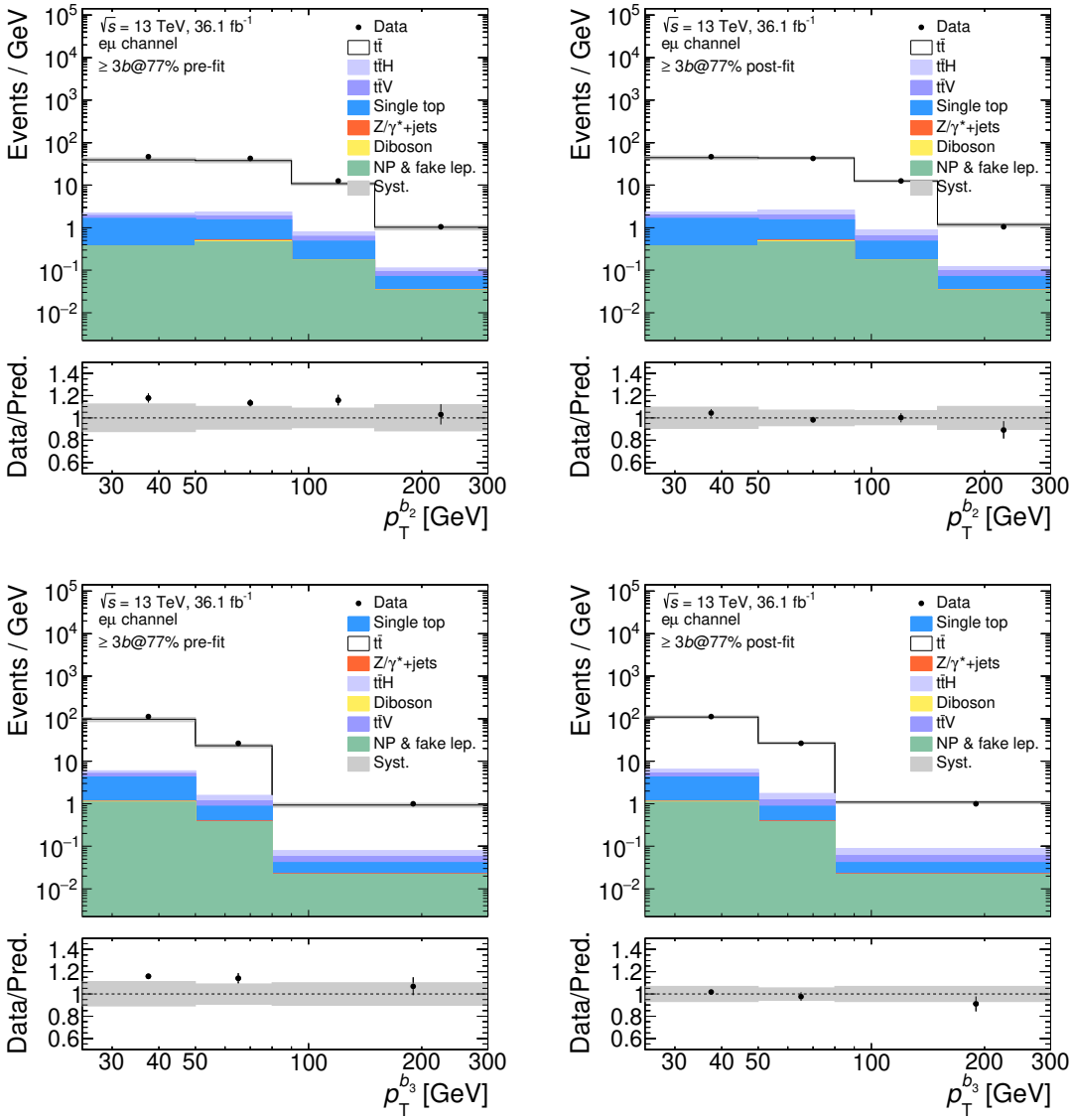


Figure 5.14.: Data to MC comparison of p_T spectra of sub-leading and third leading b-tagged jets for events fulfilling the 2b baseline selection and having at least 3 b-tagged jets. Distributions before (left) and after (right) applying the scale factors to the $t\bar{t}$ MC obtained by the flavour fit. The systematic uncertainty band (grey) includes all uncertainties from experimental sources.

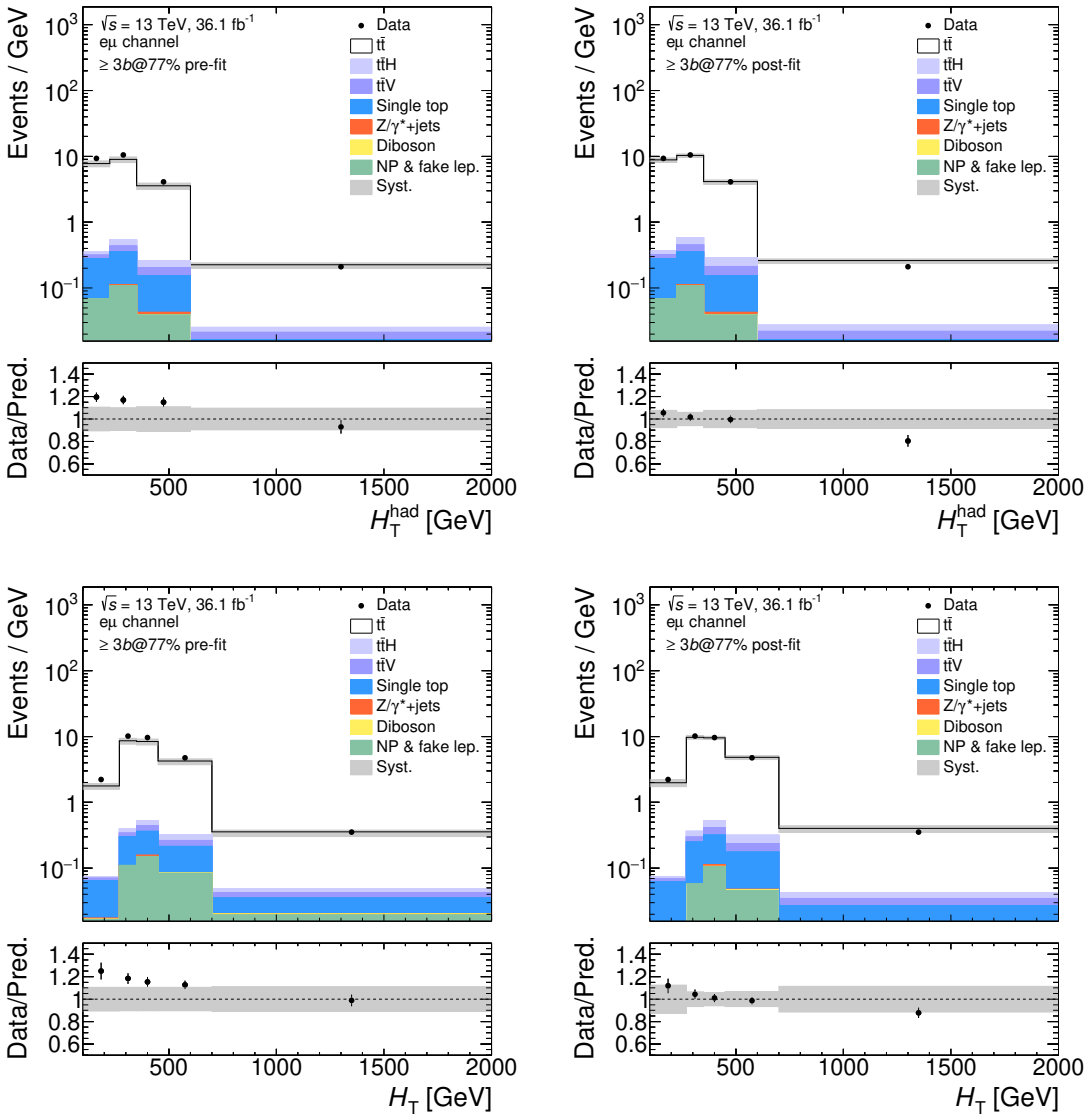


Figure 5.15.: Data to MC comparison of $H_{\mathrm{T}}^{\mathrm{had}}$ and H_{T} distributions for events having at least 3 b-tagged jets. Distributions before (left) and after (right) applying the scale factors to the $t\bar{t}$ MC obtained by the flavour fit. The systematic uncertainty band (grey) includes all uncertainties from experimental sources.

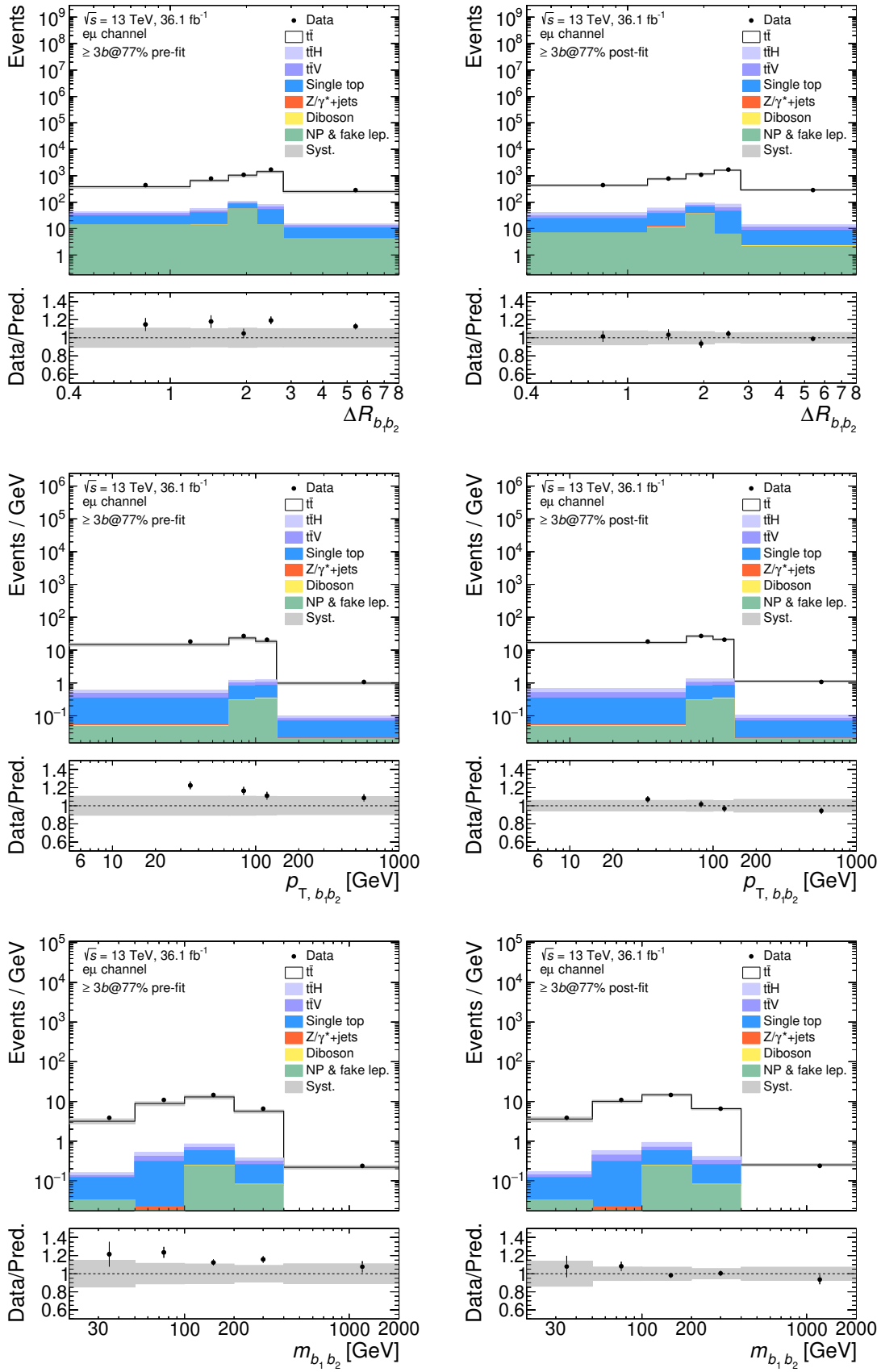


Figure 5.16.: Data to MC comparison of $\Delta R_{b_1 b_2}$, $p_{T,b_1 b_2}$ and $m_{b_1 b_2}$ distributions of two leading b-tagged jets for events having at least 3 b-tagged jets. Distributions before (left) and after (right) applying the scale factors to the $t\bar{t}$ MC obtained by the flavour fit. The systematic uncertainty band (grey) includes all uncertainties from experimental sources.

Chapter 6.

Cross-section Measurements

In this chapter the methods employed to measure the fiducial inclusive and differential cross-sections at particle-level are described. The detector-level distributions are presented in Chapter 5 and a good data-MC agreement was observed. However, no conclusions are drawn on the MC model compatibility from the detector-level data. It is to show that the predictions are close enough to data to be usable for unfolding to particle-level.

This chapter is dedicated to explain the ingredients required to perform the final cross-section measurements. The cross-section is measured at particle-level in fiducial phase space as explained in Sections 6.1. To remove the detector thumbprints and facilitate the comparison of results with other detectors or with MC simulated samples, particle-level measurements are performed. Particle-level objects are defined using stable particles before their propagation through the detector simulation. The measurements are performed in fiducial phase space which is very close to the detector-level phase space without taking into account the details of detector geometry. Section 6.2 explains the procedure to unfold the data from detector-level to particle-level and also provides the definitions of cross-sections. Section 6.3 provides the details of the tests performed to verify the stability of the procedure. The systematic uncertainty from various sources affecting the cross-section measurement are mentioned throughout the thesis in corresponding sections. Section 6.4 summarizes all these uncertainties and explains the technique used to propagate their effect in the measurement.

6.1. Particle-level Selection to Define Fiducial Phase Space

A fiducial phase-space is the sub-set of total phase space in which the measurement is performed. It is kept very close to the detector-level phase space in order to minimize the extrapolation from the measurable phase space to experimentally invisible phase space and the associated modelling dependence. The fiducial phase space is defined using particle-level

Object definitions for particle-level event selection	
Lepton selection	
Leptons	$e\mu$ pair (exactly 1 e and 1 μ with opposite charge)
Transverse momentum ${p_{\mathrm{T}}}^{e,\mu}$	$> 27\,\mathrm{GeV}$
Pseudorapidity $\lvert\eta_{e,\mu}\rvert$	< 2.5
Jet Selection	
Identification of b-jets	Ghost-matched b-hadrons with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$
Number of b-jets $N_{\text{b-jets}}$	≥ 2
Transverse momentum ${p_{\mathrm{T}}}^{b}$	$> 25\,\mathrm{GeV}$
Pseudorapidity $\lvert\eta_{b}\rvert$	< 2.5
Additional requirements	
No jet-electron or jet-muon pair with $\Delta R < 0.4$	
≥ 3 or ≥ 4 b-jets for fiducial inclusive and differential cross-sections	

Table 6.1.: Particle-level event selection in the fiducial phase space.

objects defined in Section 4.8 with the kinematic requirements similar to those placed on detector-level objects in the event selection 5.2.

The fiducial phase-space is defined by requiring exactly one electron and one muon with opposite-sign charge at particle-level and at least 2b-jets. The dressed electrons and muons are required to have $p_{\mathrm{T}} > 27$ GeV and $|\eta| < 2.5$. Experimental details of the detector like the sensitivity for electrons in the overlap between barrel and endcap are not included in the fiducial definition. The jets should have $p_{\mathrm{T}} > 25\,\mathrm{GeV}$ and $|\eta| < 2.5$. b-jets are identified by requiring that at least one b-hadron with $p_{\mathrm{T}} > 5\,\mathrm{GeV}$ is ghost associated to the jet. To ensure the consistency with detector-level selection, overlap removal between the leptons and jets are applied with $\Delta R\,(\ell, j) < 0.4$.

The fiducial inclusive cross-section is determined in a further constrained phase space by requiring one opposite-sign $e\mu$ pair and at least three (four) b-jets for the cross-section with one (two) additional b-jet (s). All the differential measurements are also performed in a phase space with at least three or four b-jets except for the distribution of number of b-jets which is measured in the fiducial volume containing the leptons and at least two b-jets. Table 6.1 summarizes the event selection at particle-level in the fiducial region.

6.2. Unfolding

The distributions are obtained at detector-level and a good data-MC agreement is observed after the application of the scale factor from the flavour fit (Figures 5.13-5.16). Now, the data distributions are unfolded to particle-level. The unfolding procedure is performed to correct for detector acceptances, efficiencies and resolution effects. The inclusive and differential cross-section measurements performed in this analysis are defined at particle-level. Using particle-level is the best compromise as compared to detector and parton level to avoid detector effects and large modelling dependencies, respectively.

An ideal, or more precisely a hypothetical, situation would be when the truth/particle-level information about an observable enters the detector, get registered by it and later can be reconstructed 100% correctly. This condition would be formulated with an equality as $R_{out} = T_{in}$, where R_{out} is the full set of information about the observable after reconstruction as output data and T_{in} is the true information at particle-level. In reality, a detector has some reconstruction limitations, non uniform efficiency, resolution effects etc. which lead to the folding/distortion of the measurement. As discussed in [170], the effect of detector response on the measurement of an observable can be expressed as:

$$R_j = \sum_i S_{ij} T_i \tag{6.1}$$

where T_i and R_j are the truth level and reconstructed distributions of an observable in bin i and j, respectively. The matrix S_{ij} represents the detector response and is referred as smearing matrix. The diagonal elements of this matrix give the probability of a truth level being correctly reconstructed, while the off-diagonal elements give the probability that a value belonging to bin i is mistakenly measured as belonging to bin j. In a measurement we obtain the reconstructed information and therefore, in order to recover the true distribution the smearing matrix S should be inverted and applied on the reconstructed distribution:

$$T_i = \sum_j S_{ij}^{-1} R_j \tag{6.2}$$

Nonetheless, it is not always possible that S_{ij} is invertible or has a unique solution. Therefore, rather than the direct inversion of the matrix, an alternative method proposed by D'Agostini called the iterative Bayesian unfolding [171] is used in this analysis.

6.2.1. Iterative Bayesian Unfolding

The iterative Bayesian unfolding technique [171] is based on the Bayes' theorem. This method can be realized in terms of several causes ($C_i, i = 1, 2, ..., n_C$) and effects ($E_j, j = 1, 2, ..., n_E$). In a scenario of experimental particle physics as used in this analysis, the cause C_i represents the number of events in each bin i of the particle-level distribution, whereas the effect corresponds to the number of reconstructed events measured from the detector in each bin j, with the total number of bins imply the total number of effects n_E. Each cause can produce different effects, but for a given effect the exact cause is not known. From the knowledge of migration and detector efficiency and resolution (like the smearing matrix S_{ij} mentioned above) obtained from MC, it is possible to estimate the probability for an effect E_j to be generated from a cause C_i, $P(E_j|C_i)$. Now the aim is to find the probability of a cause C_i which was responsible for the measured effect E_j. Using the Bayes' theorem, this probability, $P(C_i|E_j)$, can be expressed as:

$$P(C_i|E_j) = \frac{P(E_j|C_i) \cdot P_0(C_i)}{\sum_{l=1}^{n_c} P(E_j|C_l) \cdot P_0(C_l)} \tag{6.3}$$

where, $P_0(C_i)$ is an initial distribution at particle-level. The number of estimated events of the cause C_i in bin i can be given as:

$$N(C_i) = \frac{1}{\varepsilon_i} \sum_{j=1}^{n_E} P(C_i|E_j) \cdot cf_j \cdot N(E_j) \tag{6.4}$$

with $N(E_j)$ as the number of events in bin j of the effect. ε_i is the efficiency that the cause C_i has an effect E_i and cf_j is a set of correction factors that accounts for the matching and acceptance losses, to be discussed in detail below (Section 6.2.2). This can be rewritten in terms of a response matrix $\mathcal{M}$:

$$N(C_i) = \frac{1}{\varepsilon_i} \sum_{j=1}^{n_E} \mathcal{M}_{ij}^{-1} \cdot cf_j \cdot N(E_j) \tag{6.5}$$

The final value of $N(C_i)$ is derived iteratively starting from the initial distribution $P_0(C_i)$. The result from the current and previous iterations are compared using χ^2 calculation. In general the number of iterations are stopped when the value of χ^2 reaches a reasonably small value, the method adopted in this analysis is similar and is described in Section 6.3.2. The number of iterations refers to the regularisation parameter of the iterative Bayesian unfolding.

6.2.2. Unfolding Implementation

The unfolding procedure is implemented on the reconstructed events to obtain the particle-level distributions. It corrects for several effects such as detector inefficiencies, resolutions and migration of events form one bin to another for reconstructed and particle-level distributions. Some examples are: events in which one or more particle-level b-jets do not pass the p_{T} threshold for reconstruction-level b-jets or when the selection efficiency for inclusive $t\bar{t}$ events changes as a function of jet multiplicity. Furthermore, additional reconstructed jets without a corresponding particle-level jet may appear due to pile-up. If a jet migrates into the fiducial volume due to an upward fluctuation caused by the p_{T} resolution or if a single particle-level jet is reconstructed as two separate jets. All these effects are taken into account within the iterative Bayesian unfolding. The measurements are corrected separately for each observable according to

$$N^{i}_{\text{unfold}} = \frac{1}{f^{i}_{\text{eff}}} \sum_{j} \mathcal{M}^{-1}_{ij}\, f^{j}_{\text{matching}}\, f^{j}_{\text{accept}}\, f^{j}_{t\bar{t}b}\, (N^{j}_{\text{data}} - N^{j}_{\text{non-}t\bar{t}\text{-bkg}}) \tag{6.6}$$

where N^{i}_{unfold} is the number of fully corrected particle-level events in bin i. All the terms represent step by step corrections underwent in the complete unfolding procedure as described in details later. First of all, the number of events for an observable in bin j is taken from the reconstruction level which is represented by (N^{j}_{data}). The number of background events are to be subtracted and in addition a factor to correct for the $t\bar{t}c$ and $t\bar{t}l$ background $(f^{j}_{t\bar{t}b})$ is multiplied which is specific to this analysis. Then the detector acceptance (f^{j}_{accept}) is applied which rejects the events that are reconstructed at detector-level but do not pass the particle-level cuts. An additional matching factor $(f^{j}_{\text{matching}})$ is implemented in this analysis to account for the p_{T} order of the jets at both the levels. Even though the kinematic range of the measurement is chosen to be the same for particle-level and reconstruction-level objects, the migration matrix $\mathcal{M}^{-1}_{ij}$ is required to check for the migrating events. The last factor corrects for the reconstruction efficiency $\frac{1}{f^{i}_{\text{eff}}}$.

The term $(N^{j}_{\text{data}} - N^{j}_{\text{non-}t\bar{t}\text{-bkg}})$ in Equation 6.6 represents the number of events in bin j of the data distribution after subtracting the non-$t\bar{t}$ background events, determined in Section 5.3. A mixture of signal and $t\bar{t}c$ and $t\bar{t}l$ background remain in this resulting set of events. The flavour composition of the signal is corrected based on the correction factors derived in Section 5.5 as follows.

Correction for $t\bar{t}$ Related Background ($f^j_{t\bar{t}b}$)

After the subtraction of non-$t\bar{t}$ background, the data are first corrected for the $t\bar{t}$ related background which comes from the flavour mis-tagged events, by applying:

$$f^j_{t\bar{t}b} = \frac{\alpha_b N^j_{t\bar{t}b,\,\mathrm{reco}}}{\alpha_b N^j_{t\bar{t}b,\,\mathrm{reco}} + \alpha_{cl}(N^j_{t\bar{t}c,\,\mathrm{reco}} + N^j_{t\bar{t}l,\,\mathrm{reco}})} \tag{6.7}$$

where α_b and α_{cl} are the scale factors computed from the flavour-fit. $N^j_{t\bar{t}b,\,\mathrm{reco}}$, $N^j_{t\bar{t}c,\,\mathrm{reco}}$ and $N^j_{t\bar{t}l,\,\mathrm{reco}}$ are the numbers of reconstructed $t\bar{t}b$, $t\bar{t}c$ and $t\bar{t}l$ events in bin j, as predicted by MC simulation, respectively.

Acceptance Correction (f^j_{accept})

The factor f^j_{accept} corrects for the fiducial acceptance, i.e. it corrects for the events that are generated outside the particle-level phase space but pass the detector-level requirements. It is the probability of an event fulfilling the reconstruction-level selection in observable bin j, N^j_{reco}, to also be within the particle-level phase space $N^j_{\mathrm{reco}\wedge\mathrm{part}}$. It is given by

$$f^j_{\mathrm{accept}} = \frac{N^j_{\mathrm{reco}\wedge\mathrm{part}}}{N^j_{\mathrm{reco}}} \tag{6.8}$$

Figure 6.1a shows the distribution of f^j_{accept} derived for events with two b-tagged jets. The acceptances for the other observables can be found in Figure 6.2.

Matching Factor (f^j_{matching})

In distributions involving p_{T} ordered jets, a correction has to be applied for events where the order of the jets at particle-level differs from the order at detector-level. This can happen due to the limited p_{T} resolution of the reconstructed (reco) jets, in particular when two jets have similar p_{T} at particle-level. The factor f^j_{matched} corrects for potential differences in the order by comparing the order of matched jets at particle-level and at detector-level in events with at least 3 b-tagged jets matched to particle-level b-jets. Objects are considered matched if the ΔR between the particle-level and detector-level object is less than 0.4. The matching factor is defined as

$$f^j_{\mathrm{matched}} = \frac{N^{j,\,\mathrm{reco}\wedge\mathrm{part}}_{\mathrm{matched}}}{N^{j,\,\mathrm{reco}\wedge\mathrm{part}}} \tag{6.9}$$

where $N^{j,\text{reco}\wedge\text{part}}_{\text{matched}}$ refers to the number of events where the order is the same at generated particle-level jets and reconstructed detector-level jets, and $N^{j,reco\wedge part}$ is the number of all events in this selection. As shown in Figure 6.2 this correction largely varies with the observable. For the leading b-jet p_{T} the order is the same and no correction seems necessary at high p_{T}, however at low p_{T} only 60% of the events have the same order in the b-jets. For the b-jet with the 3$^{\text{rd}}$ highest p_{T}, 70% of the events have the same order independent of p_{T}. A possible explanation is that if the 3$^{\text{rd}}$ jet has a high p_{T}, the leading 2 jets probably have only a bit higher p_{T}, so they are all close in p_{T} and a change of order can happen due to energy resolution.

Response Matrix ($\mathcal{M}^{-1}_{ij}$)

The next step in the unfolding procedure is obtaining the response matrix, $\mathcal{M}^{-1}_{ij}$ from the iterative Bayesian unfolding technique (Section 6.2.1), as implemented in the RooUnfold software package [172]. It represents the probability, $P(N^{i}_{\text{part}}|N^{j}_{\text{reco}})$ of an event at particle-level in bin i to be reconstructed in bin j. The response matrices are constructed from events passing both reco and particle-level selection in fiducial phase space and where b-tagged reco jets are matched to particle-level b-jets. The binning for each observable is optimised such that the response matrix has a large fraction of events on the diagonal with sufficient number of events in each bin (Section 6.3.1). As part of the Bayesian unfolding using Equation 6.5, $\mathcal{M}^{-1}_{ij}$ is calculated iteratively, the corrected spectra are found to converge after five iterations of the Bayesian unfolding algorithm (Section 6.3.2).

Figure 6.1b shows the matrix for unfolding the b-jet multiplicity. The limited b-tagging efficiency of 77% leads to large migrations. A bit more than half of the events with 3 b-jets have only two reconstructed b-tagged jets. However, 45% still have 3 b-tagged jets. Only roughly one third of the events with 4 b-jets at particle-level are tagged as such on reconstruction level. There are no events with more b-jets at reconstruction level than at generated particle-level which is because of the matching performed in the previous step.

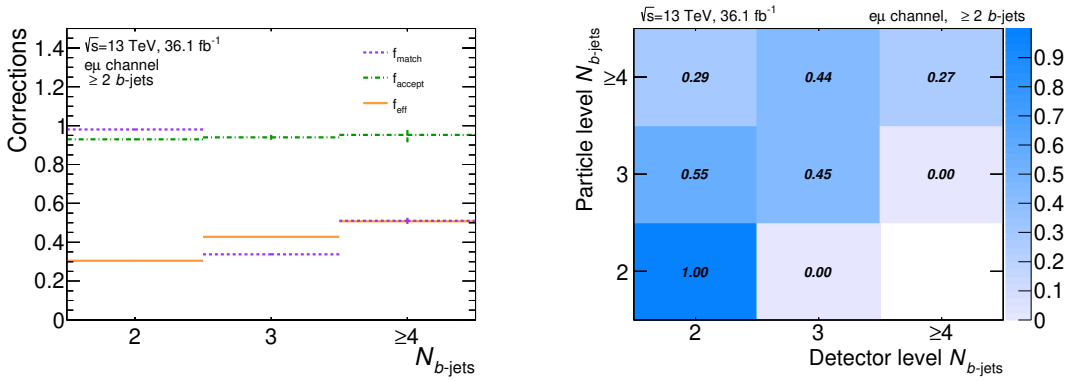
Efficiency Correction (f^{i}_{eff})

The term f^{i}_{eff} corrects for the reconstruction efficiency. It represents the efficiency to reconstruct an event in bin i, defined as the ratio of events in bin i that fulfil both the fiducial volume selection at particle-level and that at detector-level, $N^{i}_{\text{reco}\wedge\text{part}}$, to the number of events that

fulfil only the particle-level selection, N^i_{part}

$$f^i_{\text{eff}} = \frac{N^i_{\text{reco}\wedge\text{part}}}{N^i_{\text{part}}} \tag{6.10}$$

For example, in the measurement of b-jet multiplicity, the efficiency f^i_{eff}, raises from ~ 0.3 for two b-jets up to ~ 0.5 for four b-jets as can be seen in Figure 6.1a. The efficiencies for the other observable distributions can be found in Figure 6.2.



(a) **(b)**

Figure 6.1.: Efficiency (f^i_{eff}), acceptance (f^i_{accept}) and matching (f^i_{matched}) corrections as a function of b-jet multiplicity are shown in left figure. The plot on the right shows the response matrix for the events with at least 2 reconstructed b-tagged jets and at least 2 particle-level b-jets.

6.2.3. Cross-section Definition

The fiducial inclusive cross-section, σ_{fid}, is determined by integrating the absolute cross section over a variable in the measurement range. The absolute cross-section is defined as:

$$\frac{d\sigma_{\text{fid}}}{dX^i} = \frac{N^i_{\text{unfold}}}{L\Delta X^i} \tag{6.11}$$

where N^i_{unfold} is the unfolded distribution in fiducial phase space in the i th bin of the variable as described in Section 6.2.2 and ΔX^i is width of variable X in bin i. $L = 36.1\,\text{fb}^{-1}$ is the total integrated luminosity of the dataset used. The σ_{fid} is given as:

$$\sigma_{\text{fid}} = \int \frac{d\sigma_{\text{fid}}}{dX} dX = \frac{\sum_i N^i_{\text{unfold}}}{L} \tag{6.12}$$

and is also used as a normalisation factor such that results are presented in terms of a relative differential cross-section as: $\frac{1}{\sigma_{\text{fid}}}\frac{d\sigma_{\text{fid}}}{dX^i}$.

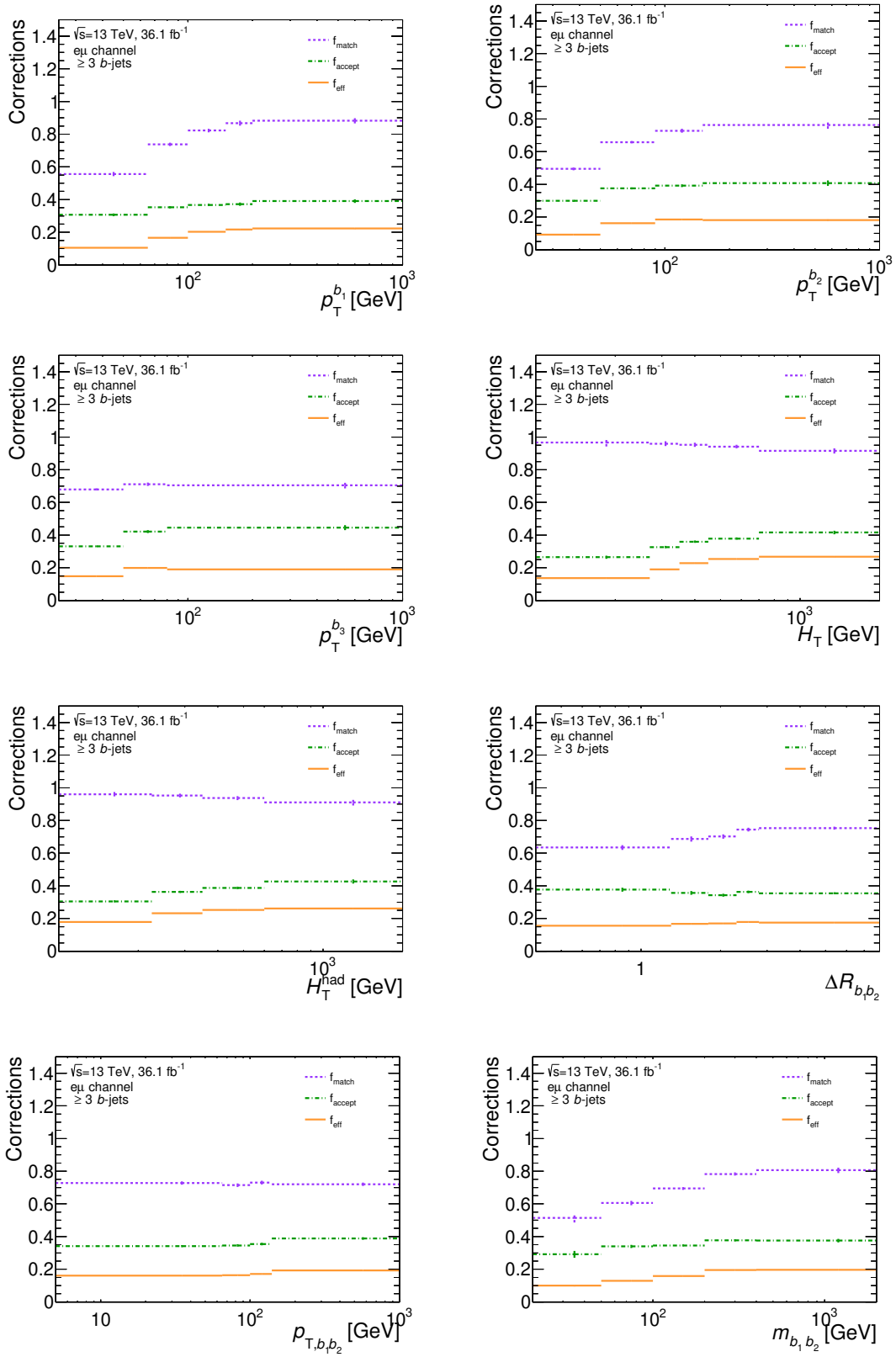


Figure 6.2.: Efficiency (f^i_{eff}), acceptance (f^i_{accept}) and matching (f^i_{matched}) corrections as a function of various observables ($p_T^{b_1}$, $p_T^{b_2}$, $p_T^{b_3}$, H_T, H_T^{had}, $\Delta R_{b_1 b_2}$, $p_{T,b_1 b_2}$, $m_{b_1 b_2}$).

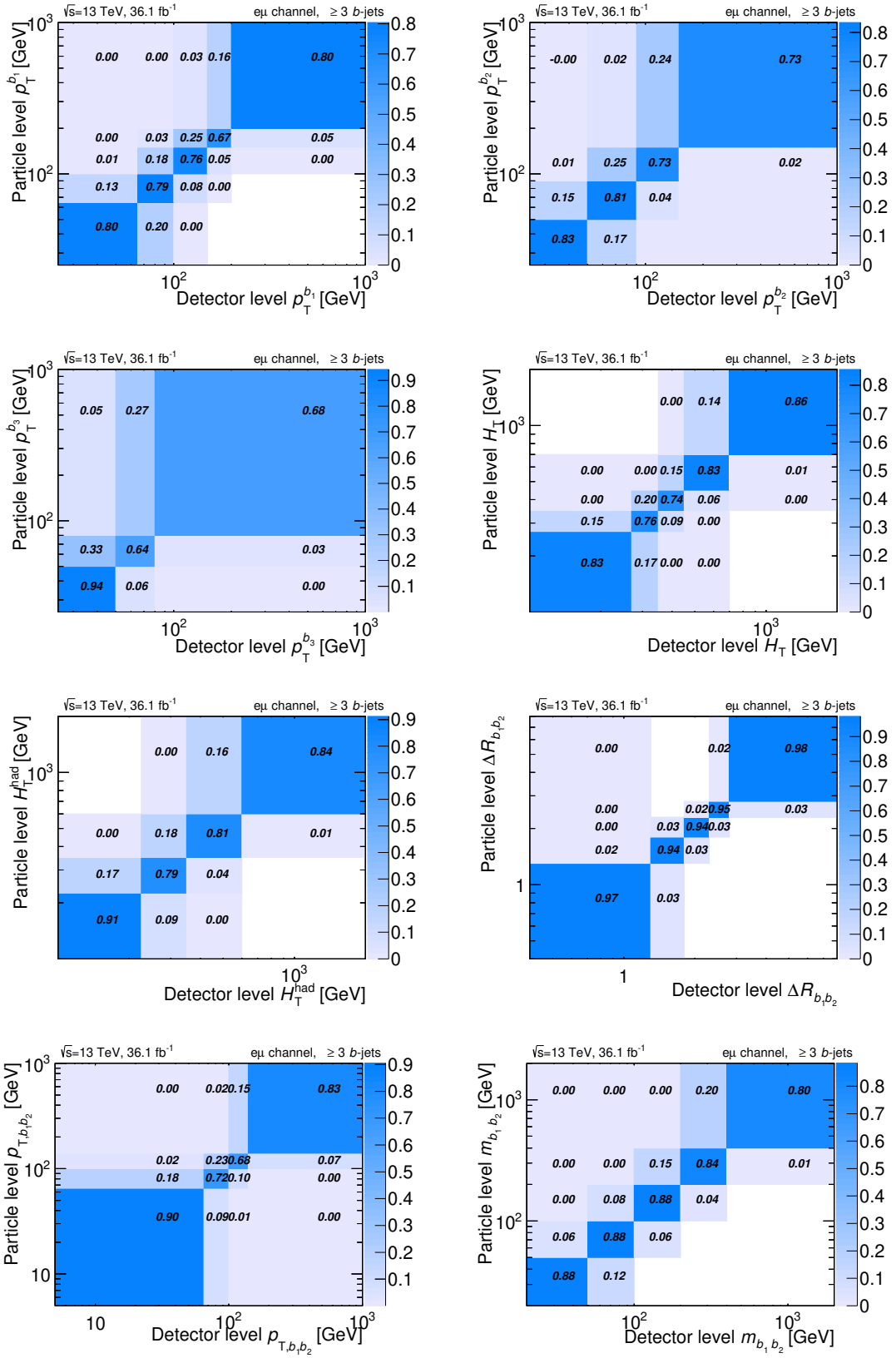


Figure 6.3.: Matrices for the detector response derived from events with at least 3 reconstructed *b*-tagged jets and at least 3 particle-level *b*-jets, and where the reconstructed and particle jet have the same order.

6.3. Unfolding Validation

The unfolding technique employed in this thesis is validated by performing several tests in order to asses the stability and reliability of the chosen method. There are two types of tests performed. One method is to test the self consistency of the procedure by using the same nominal MC sample (POWHEG + PYTHIA8) which is used to perform the unfolding in the analysis. Here, the same sample is divided into two halves and the unfolding is performed to check the stability of the procedure. These consist of pull test, optimisation of binning and optimisation of the number of iterations. It is also used to justify the choice of binning and number of iterations used. The second method is to test that the unfolding procedure can unfold distributions other than the nominal MC sample. This is referred to as *stress test* in the following.

6.3.1. Pull Test and Optimisation of binning

A test called pull-test is performed to check the presence of any bias introduced by unfolding as well as to establish a suitable choice of binning. For this study, the nominal $t\bar{t}$ MC sample, POWHEG + PYTHIA8, is divided randomly into two halves where one half is used as the *test* sample and the other half is used for the *training* sample. This procedure ensures that the test sample is statistically independent of the training sample. The test sample is used as pseudo-data to construct the detector-level and particle-level test distributions while the training sample is used to construct the unfolding corrections and response matrices. The pseudo-data is then unfolded with the corrections derived from the training sample and the pull per bin is calculated.

$$\text{pull}_i = \frac{N_i^{\text{unfold}} - N_i^{\text{truth}}}{\sigma_i^{\text{unfold}}} \tag{6.13}$$

where N_i^{unfold} is the number of entries after unfolding the distribution in bin i and N_i^{truth} is the number of entries in bin i at particle-level. σ_i^{unfold} is the statistical error on the unfolded events in the i^{th}-bin. 500 pseudo-experiments are performed and the pull distributions obtained from all pseudo-experiments are fitted to a Gaussian. If the fitted pull distribution results in a mean of zero and width one, then the measurement and statistical uncertainty of the unfolded distributions are treated correctly by the unfolding.

In addition, the bins are initially selected such that 60% of the generated particle-level events are reconstructed in the same bin, i.e 60% on the diagonal of the response matrix. Furthermore, this binning choice must result in pulls. The fit results for the pulls are plotted in terms of

bins as shown in Figure 6.4. The pull distributions in Figure 6.4 shows that the means are

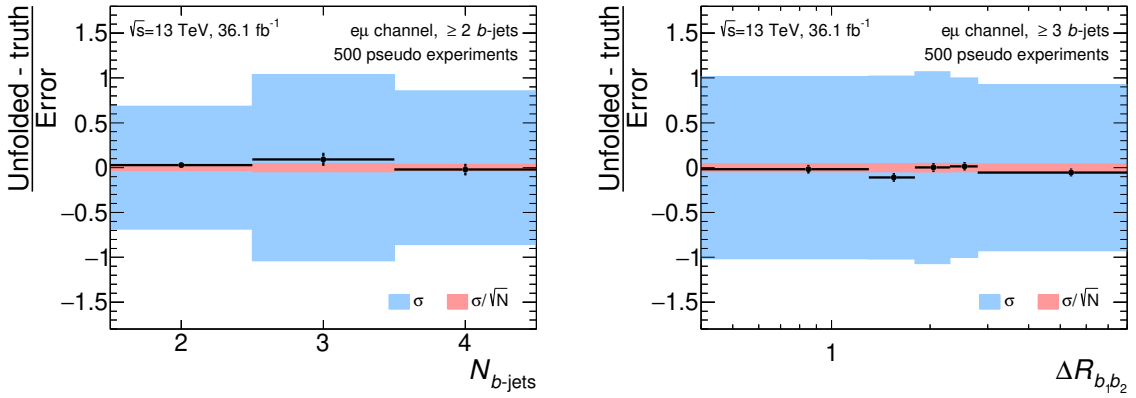
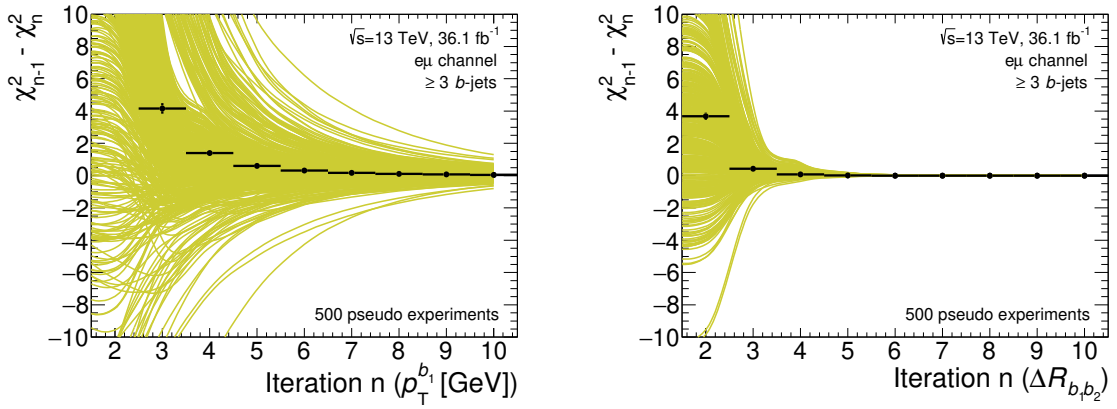


Figure 6.4.: The results of the pull tests, defined in Equation 6.13, performed with pseudo-data for (a) number of b-jets and (b) p_{T} of the third leading b-jet.

consistent with zero indicating no bias in the measurement and the widths are consistent with one showing the correct assessment of the statistical uncertainties. Since the pull is performed per bin and the same result is seen in each bin, it justifies the proper choice of binning. The test has been performed for all the distributions and are archived in Appendix A.

6.3.2. Optimisation of Number of Iterations

As the unfolding is performed using a Bayesian approach, the number of iterations is the unfolding regularisation parameter which should be optimised such that one gets stable result. In this study as well, the nominal $t\bar{t}$ sample is randomly split into two halves, one representing the test sample and the other training sample. The pseudo-data are constructed from testing sample and unfolded with corrections derived from the training samples with a given number of iterations ranging from 1 to 10. The χ^2 is computed for each iteration using the difference in the unfolded distribution with respect to the generated particle-level distribution from the test sample and by taking into account the covariance matrix of the unfolded distributions. The change in χ^2 of the n^{th} iteration and the $(\mathrm{n}-1)^{\mathrm{th}}$ iteration is compared for all iterations. This procedure is repeated 200 times. Figure 6.5 shows the distributions for the change in χ^2 with respect to the iteration for each pseudo-experiment in light green and for the mean of all pseudo-experiments in black for some variables. The remaining variables are listed in Appendix A. For most of variables the change in χ^2 is negligible after five iterations. Therefore, for consistency five iterations are used for all variables.



(a) **(b)**

Figure 6.5.: Change in χ^2 after nth iteration of Bayesian unfolding in various pseudo-experiments is shown in yellow for (a) p_T of leading b-jet and (b) ΔR between the two leading b-jets. The χ^2 test is performed between unfolded and generated particle-level distribution taking into account the covariance matrix from unfolding. The black lines shows the mean change in χ^2 which is negligible after five iterations.

6.3.3. Stress tests

The test for the stability of unfolding procedure is performed using the same nominal $t\bar{t}$ sample as the test and training samples by diving statistically into two equal parts. Now, it is required to check that the unfolding procedure does not produce any strong bias in the distribution towards the shapes corresponding to the nominal MC sample. One approach can be to introduce some Gaussian or linear re-weighting functions in order to change the shape of the distributions and perform the stress test defined below. The method used in this analysis takes alternative MC models as test sample with the nominal POWHEG + PYTHIA8 as training sample. This method tests the model independence of the unfolding.

Three variations of the alternative MC modelling is used: POWHEG + PYTHIA8 (RadHi) to account for different level of radiation, POWHEG + HERWIG7 for difference in parton shower modelling and MG5_aMC@NLO + PYTHIA8 with alternative ME generator. To perform this test the nominal POWHEG + PYTHIA8 sample is re-weighted to resemble an alternative MC sample, both at detector and particle-level, in order to change the shapes of the distributions. The resulting alternative spectrum is the test sample used to construct the pseudo-data. These pseudo-data are obtained by Poisson fluctuation of each bin of the resulting distributions. 1000 pseudo-experiments are performed where the detector-level distribution from each pseudo-data sample is unfolded with the original (un-weighted) nominal POWHEG+PYTHIA 8. The relative difference in the unfolded spectra with respect to re-weighted particle-level spectra is

computed quantified by

$$\text{bias}_i = \frac{N_i^{\text{unfold}} - N_i^{\text{truth}}}{N_i^{\text{truth}}} \tag{6.14}$$

where N_i^{unfold} is the number of entries obtained in each bin i for the unfolded spectra and N_i^{truth} is the number of entries in bin i at particle-level of the generated re-weighted spectra.

The distribution of relative differences computed from each of the 1000 pseudo-data is fit to a Gaussian with mean of zero. The fitted mean and width are plotted as a function of variable bins for some variables are shown in Figure 6.6 and the remaining are backed up in Appendix A. The mean value of the fit represents the bias for the particular MC generator, while the width of the fitted Gaussian gives an estimate for the error. The mean values of relative shifts after unfolding (black) are closer to zero within uncertainties (yellow bands). It lies well inside the statistical errors and are much smaller than the shift in the generated particle-level distributions between POWHEG+PYTHIA 8 and alternative MC (red).

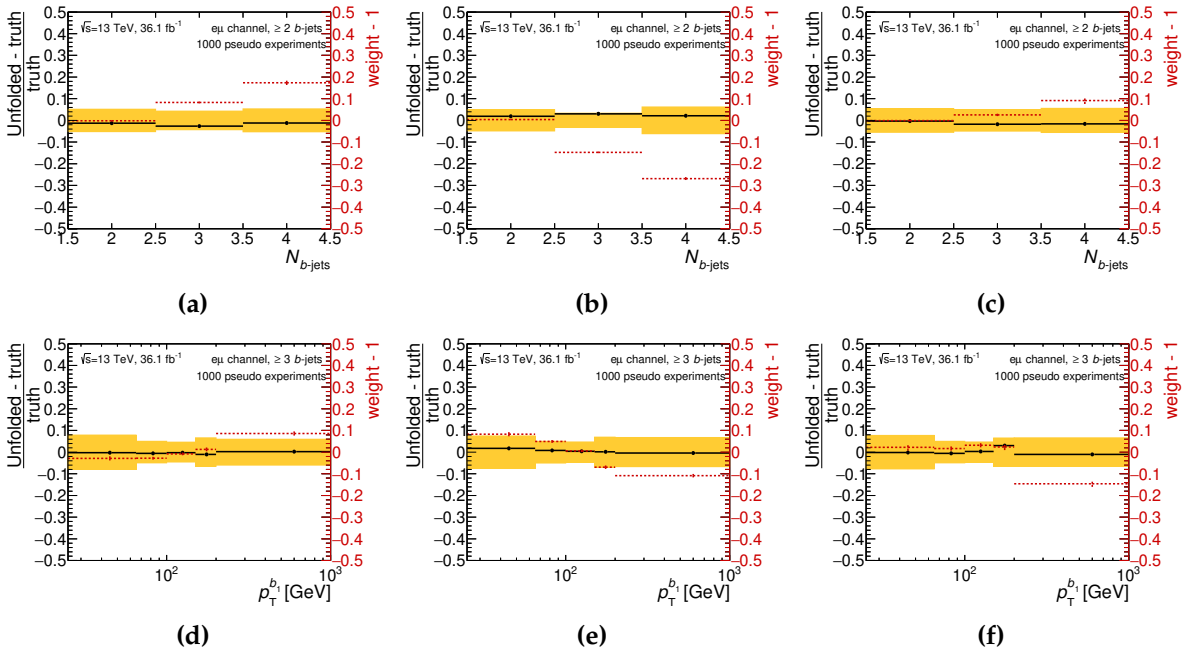


Figure 6.6.: Results showing the bias distribution for the stress test for (a-c) number of b-jets and (e-g) p_{T} of leading b-jet. The input distributions of POWHEG+PYTHIA 8 are re-weighted to POWHEG+PYTHIA 8 (RadHi) (left) to POWHEG+HERWIG7 (middle) and MG5_aMC@NLO+PYTHIA8 (right). The red dotted lines show the relative difference in shapes of the generated particle-level distributions between POWHEG+PYTHIA 8 and alternative MC, while the black lines show the relative bias after unfolding.

6.4. Systematic Uncertainty Propagation

Various sources of uncertainties, as mentioned throughout the thesis in relevant sections, affect the measurements of inclusive and differential cross-sections. This section recapitulates all these uncertainties and explains the method used to propagate their effects in the measurement. Different uncertainties need to be treated in different ways in the unfolding procedure.

6.4.1. Statistical Uncertainties

The statistical uncertainties on the unfolded data distributions are evaluated by performing 10000 pseudo-experiments. The bin contents of the data distributions before unfolding, i.e. at detector-level, are varied within the statistical uncertainty of the data which is taken to be $\sqrt{N_i}$, where N_i is the number of observed events in bin i of the distribution. To vary the content of each bin, the entry of every event is given a random weight drawn from a Poisson distribution with a mean of N_i. A replica of 10000 histograms is created in this way for the data at detector-level. Each of these histograms is then unfolded using the unfolding procedure described in Section 6.2. The unfolding corrections are derived from the nominal MC and the correction for $t\bar{t}$ related background is not evaluated separately for each pseudo-experiment but rather fixed as described in Section 5.5. The standard deviation of the spread of each bin content across all unfolded histogram replicas is then taken as the measure of statistical uncertainty in that bin.

Multiplying the event weights with random Poisson fluctuation ensures that the correlation between bins of different distributions are conserved. Figure 6.7 presents the correlation matrices for the statistical uncertainty showing the perfect bin-by-bin correlation of the variables. Rest of the variables are archived in Appendix B.

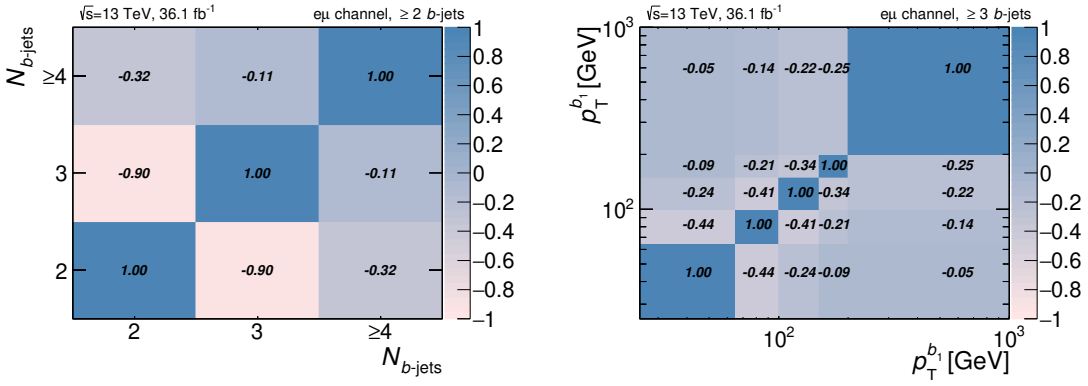


Figure 6.7.: Correlation matrices obtained from the statistical uncertainty of data evaluated on (a) b-jet multiplicity and (b) p_{T} of leading b-jet normalised distribution.

6.4.2. Experimental Uncertainties

The experimental uncertainties specifically refer to the systematic uncertainties related to detector performance such as detector resolution and object reconstruction. However, this section also lists the $t\bar{t}$ and non-$t\bar{t}$ background related uncertainties because the method to propagate these uncertainties are the same. Hence, the experimental uncertainties influencing the unfolded results are:

Detector systematic uncertainties: These sources include the integrated luminosity, pile-up and systematic uncertainties related to object reconstruction, specifically leptons and jets. The luminosity and pile-up uncertainties are quoted in Section 2.3. As mentioned in Chapter 4, object reconstruction, identification, trigger efficiency, energy scale and energy resolutions are derived in MC simulation and are corrected with scale factors to match the data. The uncertainties on these scale factors are to be propagated to the measurements. Most of these uncertainties are determined by the dedicated performance groups and are provided centrally for all physics analyses.

The detector systematic uncertainties include: the electron and muon momentum scale and resolution, the jet energy scale and resolution consisting of JER and 21 components of JES, JVT calibration, the b-jet identification with 30 components related to the b-tagging efficiencies and 15(80) components related to the mis-tag rates of c-jets (light-flavour jets). Each systematics component is taken uncorrelated with respect to the other components.

$t\bar{t}$ and non-$t\bar{t}$ related uncertainties: The uncertainty originating from the $t\bar{t}$ background is determined by changing the fraction of c- and light-jets in the sample and is explained in Section 5.5. The non-$t\bar{t}$ background is composed of non-prompt or fake leptons, Drell-Yan, single-top and diboson processes. The uncertainties considered on these background processes are mentioned in Section 5.3 along with their estimations.

The procedure used to propagate the experimental uncertainties is an extension of what is used for the statistical uncertainty. 10000 pseudo experiments are generated by Poisson fluctuating the bin contents of the detector-level kinematic distributions observed in data. To introduce the effect of experimental uncertainty, each pseudo-data is smeared using Gaussian distribution.

$$N_i^{\text{toy}} = P^{\text{toy}}(N_i)\left(1 + \sum_{k=1}^{n_{\text{syst}}} \delta_{k,i}\,\xi_k^{\text{toy}}\right) \tag{6.15}$$

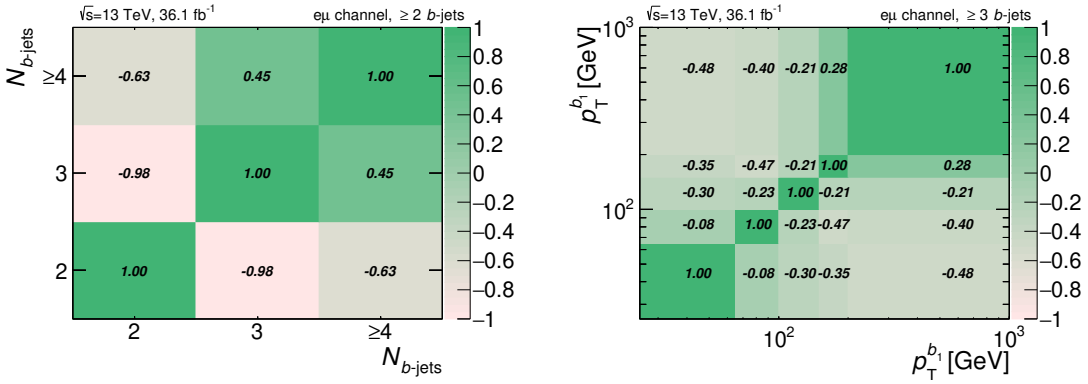
where $P^{\text{toy}}(N_i)$ represents the Poisson distribution with mean N_i, the number of observed data events in i^{th} bin of the distribution of the variable. The index k refers to all the experimental systematic uncertainties mentioned above. For each systematic uncertainty, the relative variation due to that uncertainty is obtained at the detector-level, using the nominal MC sample. $\delta_{k,i}$ is the relative systematic uncertainty for component k in bin i, and ξ_k^{toy} is a random number extracted from a Gaussian with mean of zero and width of one.

Finally, the N_i^{toy} spectra obtained from Equation 6.15 are required to be unfolded. To determine the data-driven $t\bar{t}$ related correction, $f_{t\bar{t}b}$ in Equation 6.6, a fit is performed for each of the systematic uncertainty to get the re-weighting factors. The correction factors are not extracted for each individual pseudo-experiment, instead obtained only once for each systematic error. The re-weighting factors for some of the important systematics are listed in Table 6.2. The acceptance, response matrix and efficiency corrections are derived from the nominal POWHEG + PYTHIA8 sample after the sample is re-weighted using the α_b and α_{cl} factors derived by fitting the POWHEG + PYTHIA8 templates to data.

Variations	α_b	α_{cl}
PowPy8 to data	1.37 ± 0.06	1.05 ± 0.04
PowPy8 to JET/JER_up	1.01 ± 0.06	1.03 ± 0.04
PowPy8 to JET/JER_down	0.98 ± 0.06	0.97 ± 0.03
PowPy8 to BTAG_*b*-jet	1.03 ± 0.06	1.02 ± 0.04
PowPy8 to BTAG_*c*-jet	1.02 ± 0.06	1.00 ± 0.04
PowPy8 to BTAG_light-jet	0.99 ± 0.06	1.12 ± 0.04
PowPy8 to Sherpa	1.31 ± 0.07	1.13 ± 0.04
PowPy8 to aMC@NLO	0.92 ± 0.06	0.96 ± 0.03
PowPy8 to PowHW7	0.69 ± 0.05	0.93 ± 0.03
PowPy8 to RadLo	0.92 ± 0.05	0.87 ± 0.03
PowPy8 to RadHi	1.04 ± 0.06	0.96 ± 0.04

Table 6.2.: Fitted scale factors α_b and α_{cl} for the systematic variation samples. The quoted uncertainties are statistical uncertainties on the fit.

The errors are then evaluated by taking the relative difference of each experiment's result with respect to the mean for 16%, 50% and 84% of the toys to determine the lower and upper bounds. The difference from the lower bound to mean, and from the mean to the upper bound are considered to give the downward and upward uncertainties on the measurements, respectively. Figure 6.8 shows the bin-by-bin correlation obtained from experimental systematic uncertainties for the variables.



(a) **(b)**

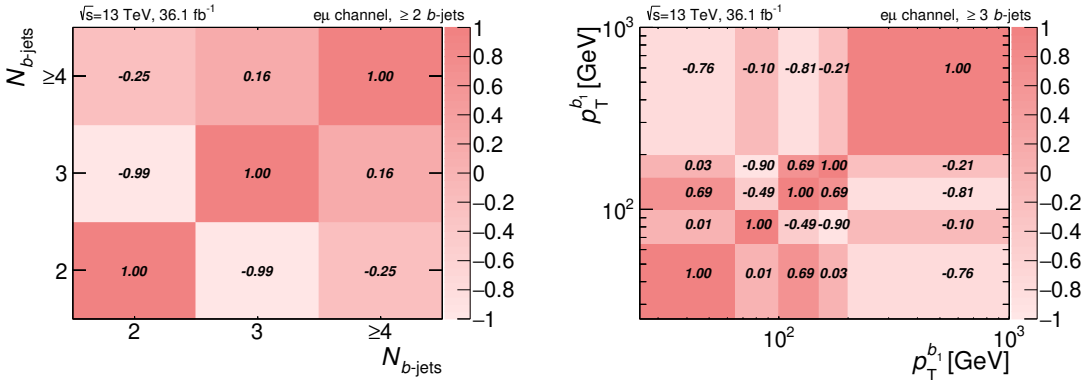
Figure 6.8.: Correlation matrices obtained from experimental systematic uncertainties evaluated on (a) b-jet multiplicity and (b) p_{T} of the leading b-jet normalised distribution.

6.4.3. Modelling Uncertainties

The modelling uncertainties are assigned to the measurement in order to cover the biases due to the choice of a particular $t\bar{t}$ MC model. Comparison between different samples can probe different aspects of the modelling and can therefore be used to assess the associated modelling uncertainties. The samples used to target a particular modelling component such as the parton shower, generator, ISR/FSR and PDFs are mentioned in Section 3.5. The uncertainties on the $t\bar{t}H$ and $t\bar{t}V$ samples are mentioned in Section 3.4.1.

To propagate the modelling uncertainties, the detector-level distributions from alternative MC samples are unfolded using the unfolding corrections derived from the nominal POWHEG + PYTHIA8 sample. The correction factor for the $t\bar{t}$ background is determined by the template fit to the corresponding MC prediction and deriving the re-weighting factors as quoted in Table 6.2. The result of this unfolding is then compared with the particle-level distribution of the alternative MC sample. The relative difference between the result and the particle-level distribution in each bin is assigned as a systematic uncertainty. The correlation between the bins are shown in Figure 6.9.

An uncertainty due to the limited size of the MC samples used is evaluated in a similar way as the evaluation of data statistics. The nominal MC prediction, POWHEG + PYTHIA8, is used as pseudo-data with each MC event given a random event weight drawn from a Poisson distribution. This is performed five thousand times to create five thousand pseudo-data distributions. Each of these pseudo-data distributions is unfolded using the nominal unfolding setup and the standard deviation in each bin is taken as the uncertainty due to limited MC statistics.



(a) **(b)**

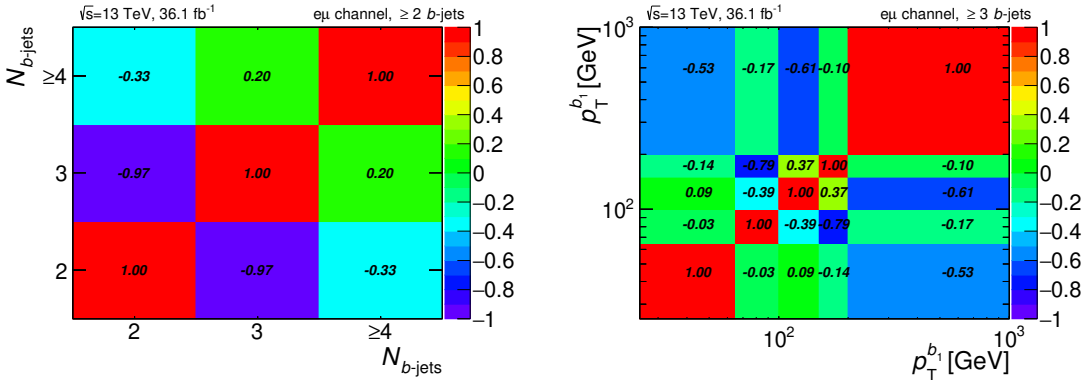
Figure 6.9.: Correlation matrices obtained from systematic uncertainties related to modelling evaluated on (a) b-jet multiplicity and (b) p_{T} of leading b-jet normalised distribution.

6.5. Total Uncertainty

The total systematic uncertainty, δ_{syst}, is obtained by taking the square root of the quadrature sum of the experimental δ_{exp} and modelling δ_{mod} uncertainties.

$$\delta_{\mathrm{syst}} = \sqrt{(\delta_{\mathrm{exp}})^2 + (\delta_{\mathrm{mod}})^2} \tag{6.16}$$

Figure 6.10 shows the correlation between the bins of the variables for the total uncertainty which includes both statistical and systematic uncertainties.



(a) **(b)**

Figure 6.10.: Correlation matrices obtained from statistical and systematic uncertainties evaluated on (a) b-jet multiplicity and (b) p_{T} of leading b-jet normalised distribution.

Figures 6.11 shows the impact of some of the major uncertainty sources on the measured particle-level spectra. The dominant uncertainties are originating from the modelling of the signal sample. The data driven template fit to correct for the mis-tagging of c- and light jets

as b-jets also incorporates relatively bigger uncertainty. The uncertainties from JES/JER and b-tagging are also significant.

Figures 6.11c and 6.11d illustrates the major contributions of the systematic uncertainties for H_{T} and $H_{\mathrm{T}}^{\mathrm{had}}$. Parton shower modelling is the dominant uncertainty in most regions of $H_{\mathrm{T}}^{\mathrm{had}}$. Similar uncertainties are found in the measurement of H_{T}, where the low H_{T} region has relatively larger uncertainties due to the QCD radiation scale variations because of softer jets contributing to this region. The contribution of different types of uncertainties on other observables are presented in Appendix B.

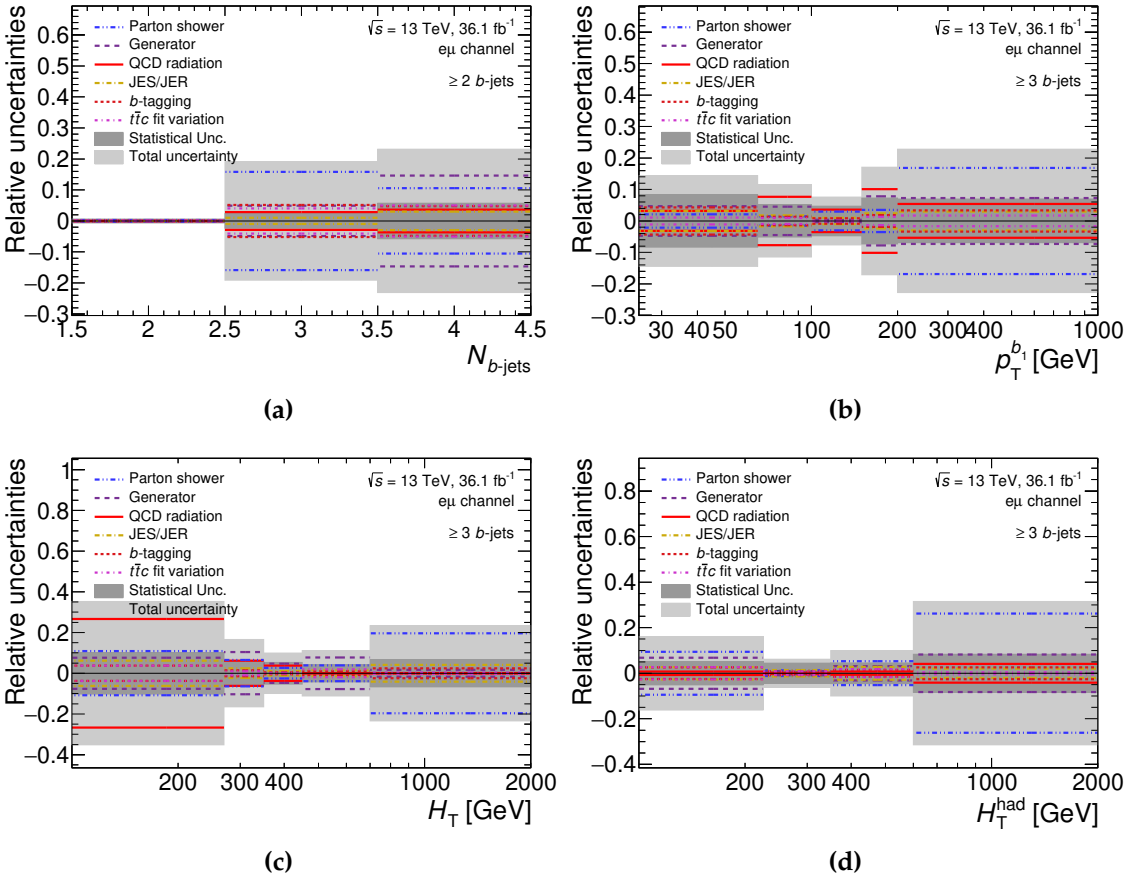


Figure 6.11.: Relative systematic uncertainties from various theoretical and experimental sources on the particle-level distributions as a function of (a) number of b-jets, (b) p_{T} of leading b-jet, (c) H_T of all the final state particles and (d) $H_{\mathrm{T}}^{\mathrm{had}}$ of all the jets. The dark band shows the statistical uncertainty while the light band shows the statistical and systematic uncertainty.

Chapter 7.

Results

This chapter presents the unfolded results as inclusive fiducial cross-section and normalised differential fiducial cross-sections. The measured fiducial cross-sections with associated uncertainties for $t\bar{t}$ production in association with at least one or at least two additional b-jets are shown in Section 7.1. The cross-section from various state-of-the-art MC predictions are reported for comparison. Section 7.2 elaborates the differential cross-sections as a function of various interesting observables and their comparison with different MC models.

7.1. Inclusive Fiducial Cross-sections

The inclusive fiducial cross section for $t\bar{t}$ production with at least one (two) additional b-jet (s) is obtained according to Equation 6.12. The quoted uncertainties are statistical and systematic uncertainties.

$$\sigma_{t\bar{t}b} = 181 \pm 5\,(\text{stat}) \pm 24\,(\text{syst})\ \text{fb} = 181 \pm 25\ \text{fb} \tag{7.1}$$

$$\sigma_{t\bar{t}b\bar{b}} = 27 \pm 3\,(\text{stat}) \pm 7\,(\text{syst})\ \text{fb} = 27 \pm 7\ \text{fb} \tag{7.2}$$

Table 7.1 shows the breakdown of uncertainties in these cross-sections. The uncertainties are dominated by the systematic uncertainty, mainly coming from the $t\bar{t}$ modelling, the data-driven template fit to correct for the mis-tagging effects (Section 5.5), b-tagging and jet related uncertainties. The uncertainties related to the jet energy scale and resolution (JES/JER) increase with increasing number of jets in the analysis as expected. Similarly, the b-tagging uncertainty increases when more jets are required to be b-tagged. It is observed that most of the modelling uncertainties related to parton shower, ISR/FSR and PDF are reduced in $t\bar{t}b\bar{b}$ ($\geq 4b$) as compared to $t\bar{t}b$ ($\geq 3b$) phase space. This implies that most of the MC models perform well for the full $t\bar{t}b\bar{b}$ process whereas modelling of $t\bar{t}b$ is less accurate. However, the generator modelling uncertainty, which is evaluated from Sherpa 2.2 $t\bar{t}$, shows the opposite trend. The

reason behind this behaviour is the description of one additional parton in the matrix element at NLO in Sherpa 2.2 $t\bar{t}$.

Fiducial Phase-space Sources	$\geq 3b$ unc. [%]	$\geq 4b$ unc. [%]
Data statistics	2.7	9.0
Luminosity	2.1	2.1
JES/JER	2.6	4.3
b-tagging	4.5	5.2
Lepton	0.9	0.8
Pile-up	2.1	3.5
$t\bar{t}c$ fit variation	5.9	11
Non-$t\bar{t}$ bkg	0.8	2.0
Detector+background total syst.	8.5	14
Parton shower	9.0	6.5
Generator	0.2	18
ISR/FSR	4.0	3.9
PDF	0.6	0.4
$t\bar{t}V/t\bar{t}H$	0.7	1.4
MC sample statistics	1.8	5.3
$t\bar{t}$ modelling total syst.	10	20
Total syst.	13	24
Total	13	26

Table 7.1.: Main systematic uncertainties in percentage for the particle-level measurement of inclusive cross-sections in the $\geq 3b$ and the $\geq 4b$ phase space.

7.1.1. Comparison with Theoretical Predictions

In order to facilitate the comparison with the theoretical calculations, the contributions mainly from $t\bar{t}H$ (Figure 1.11b) and $t\bar{t}Z$ (Figure 1.11c) are subtracted from the signal. The contributions from heavy boson production in association with $t\bar{t}$ are together referred to as $t\bar{t}X$. The subtraction is based on the predictions obtained from simulation [124] using MG5_aMC@NLO +PYTHIA8. The inclusive fiducial cross-sections after subtracting the $t\bar{t}X$ contribution are

$$\sigma_{t\bar{t}b}^{\text{Data}-t\bar{t}X} = 177 \pm 5\,(\text{stat}) \pm 24\,(\text{syst})\ \text{fb} \tag{7.3}$$

$$\sigma_{t\bar{t}b\bar{b}}^{\text{Data}-t\bar{t}X} = 25 \pm 3\,(\text{stat}) \pm 7\,(\text{syst})\ \text{fb} \tag{7.4}$$

The measured fiducial cross-section for the QCD production of $t\bar{t}b$ and $t\bar{t}b\bar{b}$ are compared to a series of theoretical predictions. These MC predictions can be divided broadly into two groups: one with NLO calculations for $pp \to t\bar{t}$ interfaced with a parton shower model (Section 3.4.2) and other with $pp \to t\bar{t}b\bar{b}$ NLO matrix element calculation interfaced with parton shower models (Section 3.4.3).

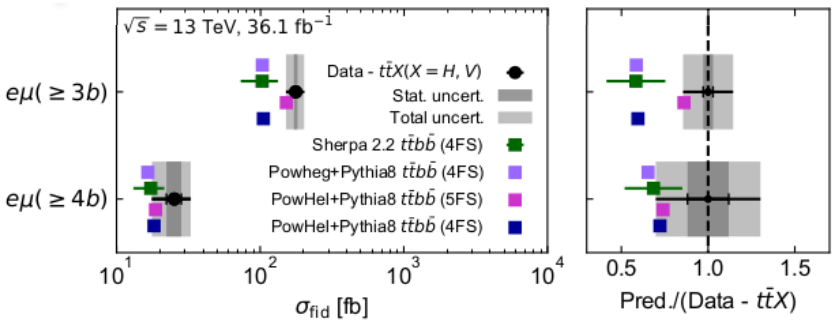


Figure 7.1.: The measured fiducial cross-sections, with $t\bar{t}H$ and $t\bar{t}V$ contributions subtracted from data, compared with $t\bar{t}b\bar{b}$ predictions obtained using SHERPA2.2 $t\bar{t}b\bar{b}$ with uncertainties obtained by varying the renormalisation and factorisation scales by factors of 0.5 and 2.0 and including PDF uncertainties. Comparisons with the central values of the predictions of POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ and POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ are also made. No uncertainties are included for the subtraction of the $t\bar{t}H$ or $t\bar{t}V$ predictions.

Figure 7.1 presents the comparison of the measurement with various $t\bar{t}b\bar{b}$ predictions. The measured cross-section for the $\geq 3b$ fiducial phase space is more precise, having an uncertainty of only 13%, than that for $\geq 4b$ phase space which has an uncertainty of 28%. The measured cross-section for the $t\bar{t}b$ process is more precise than the theoretical predictions having an uncertainty of 20%-30% [64]. The measurement exceeds the predictions in both the phase spaces but lies within the total uncertainty for the $t\bar{t}b\bar{b}$ process. This effect is larger in the $t\bar{t}b$ process but still lies within 2σ. These predictions have full $t\bar{t}b\bar{b}$ process in ME which implies that these models describe the data well when both the b-jets are in the detector acceptance in the measurement. It can be inferred that the events containing small angle gluon splitting or b-jets with low p_{T} might not be modelled very well in these MC samples and therefore the discrepancy is larger in the $\geq 3b$ phase space. It should be noted that the POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS) predictions are very recent and require some fixes before any conclusion is made regarding that. It has already been reported to the concerned team and for this reason this MC model is not discussed repeatedly in the differential cross-section results.

The comparisons are also made with the $t\bar{t}$ predictions as shown in Table 7.2. It can be seen that these predictions are closer to the measurement for the $\geq 3b$ phase space as compared to the $t\bar{t}b\bar{b}$ predictions which implies that the softer and collinear effects are well described by the parton shower models. The Sherpa 2.2 $t\bar{t}$ model fits the measurement very well for $t\bar{t}b$ phase space as well because it consists of one additional jet in the ME calculation.

Fiducial Phase-space Measurement and Predictions	$\geq 3b$ [fb]	$\geq 4b$ [fb]
Data	181 ± 5 (stat) ± 24 (syst)	27 ± 3 (stat) ± 7 (syst)
$t\bar{t}X(X = H, V)$ MC	4	2
Data $- t\bar{t}X$	177	25
$t\bar{t}$ predictions		
POWHEG+PYTHIA 8	133 ± 0.3	20 ± 0.1
POWHEG+PYTHIA 8 (RadLo)	122 ± 0.3	18 ± 0.1
POWHEG+PYTHIA 8 (RadHi)	144 ± 0.4	23 ± 0.2
POWHEG+HERWIG7	111 ± 0.3	15 ± 0.1
MG5_aMC@NLO+PYTHIA8	134 ± 0.6	21 ± 0.3
$t\bar{t}$ +jets predictions		
Sherpa 2.2$t\bar{t}$	172 ± 0.9	24 ± 0.3
$t\bar{t}b\bar{b}$ predictions		
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	103 ± 30	17.3 ± 4.2
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	104	16.5
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	152	18.7
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	105	18.2

Table 7.2.: Predicted fiducial cross-section for the phase spaces $\geq 3b$ and $\geq 4b$ compared with the $t\bar{t}X$ subtracted data. The first list with $t\bar{t}$ MC predictions has the quoted errors due to MC statistics. For SHERPA2.2 $t\bar{t}b\bar{b}$ the theoretical uncertainty is quoted.

7.2. Normalised Differential Fiducial Cross-sections

In this section the results for normalised differential cross-sections at particle-level in the fiducial phase space are presented and compared to various MC predictions. The fiducial phase space having at least one additional b-jets i.e. $t\bar{t}b$ process is presented and discussed in detail. However, the $t\bar{t}b\bar{b}$ process with at least two additional b-jets (i.e. at least 4 b-jets in total) has very low statistics, therefore it is discussed briefly in the end and all the results are documented in Appendix D. The differential distributions are presented as a function of various observables mentioned in Section 5.6.

Before proceeding to the results, a brief description of the method used to perform a quantitative test for the comparison of differential measurement with various theoretical predictions is provided. The measurement for each observable is compared with the predictions using χ^2 tests. The covariance matrices for statistical only and experimental systematics are produced

using 10000 pseudo experiments of detector-level distributions and unfolding them. Whereas, the modelling systematics are evaluated using post-unfolded non-closure with respect to particle-level as explained in Section 6.4. Thereafter, the full covariance matrices for the total (statistical and systematic) uncertainties are described and shown in Section 6.5. The χ^2 is determined for the normalised distributions using:

$$\chi^2 = S^T_{b-1}\mathrm{Cov}^{-1}_{b-1}S_{b-1} \tag{7.5}$$

where $b-1$ shows one less than the total number of bins for each distribution. Since the normalised distributions are used, one bin is discarded in the calculation to reflect the normalization constraint. S_{b-1} is a column vector representing the difference between the unfolded data and the MC generator predictions of the normalised cross-section. Cov_{b-1} is a matrix with $b-1$ rows and the respective $b-1$ columns of the full covariance matrix. The full covariance matrix is singular and non-invertible, as it is evaluated using normalised distributions. The resulting value of χ^2 is converted into p-values. The p-values are determined using the number of degrees of freedom (ndof) for each distribution. The number of degrees of freedom is defined as the number of bins minus one in case of the normalised differential cross-section. The χ^2 and p-values are tabulated for all the observables.

7.2.1. b-jet Multiplicity

Figure 7.2 shows the b-jet multiplicity spectrum. All MC predictions are 20 - 40% below the data for events with 3 b-jets and more than 40% below for events with at least 4 b-jets. The resulting χ^2 values are shown in Table 7.3, where the second column is for the normalised b-jets multiplicity distribution with $N_{b\text{-jets}} \geq 2$ and the last column is for the normalised b-jets multiplicity distribution with $N_{b\text{-jets}} \geq 3$. All MC predictions that calculate the top-quark pair production matrix element at NLO, but rely on the parton shower for high jet multiplicities, predict too few events with three or four b-jets. This suggests that the b-jet production by the parton shower is not optimal in these MC models. In order to probe the sensitivity of the process to the choice of the scale, the renormalisation and factorisation scales in the matrix element calculation and in the parton shower are changed by factors of 0.5 and 2. The situation does not improve significantly as shown in the middle ratio panel of Figure 7.2. SHERPA2.2 $t\bar{t}$, which models one additional-parton process at NLO accuracy and up to four additional partons at LO accuracy, is the only one of the presented generators that describes the b-jet production well over the full phase space.

Predictions that include additional massive b-quarks in the matrix element calculation such as: SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS), POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS), POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS) do not

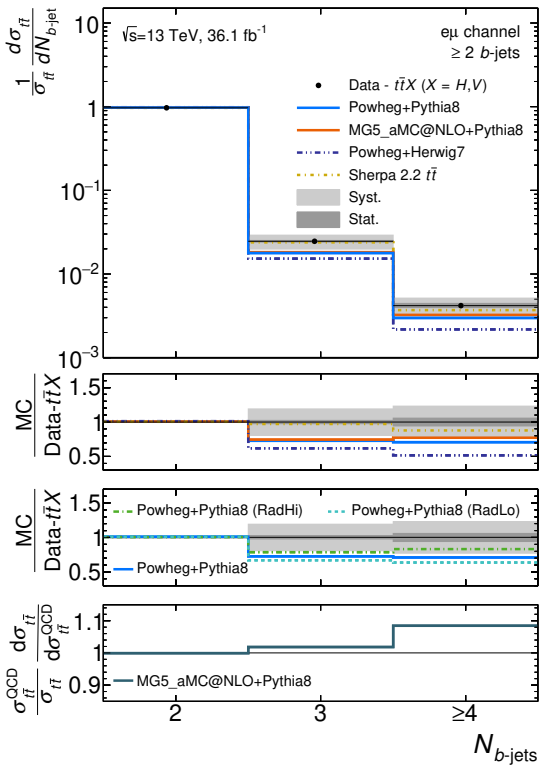


Figure 7.2.: Normalised fiducial cross sections as a function of b-jet multiplicity in events with at least 2 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions.

Generators	$N_{b-\text{jets}}$			
	$\geq 2b$		$\geq 3b$	
	χ^2 / ndof	p-value	χ^2 / ndof	p-value
POWHEG+PYTHIA 8	18.1 / 2	< 0.01	< 0.01 / 1	1.0
MG5_aMC@NLO+PYTHIA8	14.1 / 2	< 0.01	0.05 / 1	0.83
SHERPA2.2 $t\bar{t}$	0.85 / 2	0.65	0.06 / 1	0.80
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	-	-	0.37 / 1	0.54
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	-	-	0.33 / 1	0.56
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	-	-	0.76 / 1	0.38
POWHEG+HERWIG7	39.4 / 2	< 0.01	0.26 / 1	0.61
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	-	-	0.28 / 1	0.60
POWHEG+PYTHIA 8 (RadHi)	9.2 / 2	0.01	0.08 / 1	0.77
POWHEG+PYTHIA 8 (RadLo)	27.0 / 2	< 0.01	0.01 / 1	0.92

Table 7.3.: Values of χ^2 per degree of freedom and p-values between the unfolded normalised cross-section and the predictions for b-jet multiplicity. The number of degrees of freedom is equal to the number of bins minus one. The χ^2/ndof and p-values for the prediction from SHERPA $ttbb$ ME is computed only for the $\geq 3b$ selection.

provide top-pair production without additional b-jets and cannot be compared with the region with less than three b-jets. Therefore, Table 7.3 includes χ^2 values where the total additional b-jet production has been adjusted through the normalisation to $N_{b\text{-jets}} \geq 3$. The relative rate of one, two and more than two additional b-jets is described well by all predictions. It is also interesting to note that parton shower generators predict the relative rate of one and two additional b-jets well once the total additional b-jet production has also been adjusted through the normalisation to $N_{b\text{-jets}} \geq 3$.

The comparison of the predictions from various MC generators with the data are made after subtracting the simulation-estimated contributions of $t\bar{t}X$ production from the data. The third ratio panel of Figure 7.2 shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions from the $t\bar{t}X$ processes. The impact of including these processes in the prediction increases with b-jet multiplicity, resulting in a change of about 10% relative to the pure QCD $t\bar{t}$ prediction in the inclusive four-b-jet bin.

7.2.2. H_{T} and $H_{\mathrm{T}}^{\mathrm{had}}$

Distributions of H_{T} and $H_{\mathrm{T}}^{\mathrm{had}}$ are shown in Figure 7.3 for the fiducial phase space ≥ 3 b-jets. The quantitative assessment of the level of agreement between the data and several MC predictions are presented in Table 7.4. The data are well described by all MC models within uncertainties of 10%-30%.

Generators	H_{T}		$H_{\mathrm{T}}^{\mathrm{had}}$	
	χ^2/ndof	p-value	χ^2/ndof	p-value
POWHEG+PYTHIA 8	0.95 / 4	0.92	2.68 / 3	0.44
MG5_aMC@NLO+PYTHIA8	3.71 / 4	0.45	3.72 / 3	0.29
SHERPA2.2 $t\bar{t}$	0.58 / 4	0.97	2.26 / 3	0.52
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	0.35 / 4	0.99	0.40 / 3	0.94
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	4.88 / 4	0.30	1.85 / 3	0.60
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	1.39 / 4	0.85	3.33 / 3	0.32
POWHEG+HERWIG7	0.26 / 4	0.99	2.28 / 3	0.52
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	0.63 / 4	0.96	3.93 / 3	0.27
POWHEG+PYTHIA 8 (RadHi)	4.09 / 4	0.39	6.43 / 3	0.09
POWHEG+PYTHIA 8 (RadLo)	0.14 / 4	1.0	1.06 / 3	0.79

Table 7.4.: Values of χ^2 per degree of freedom and p-values between the unfolded normalised cross-section and the predictions for H_{T} and $H_{\mathrm{T}}^{\mathrm{had}}$ measurements. The number of degrees of freedom is equal to the number of bins minus one.

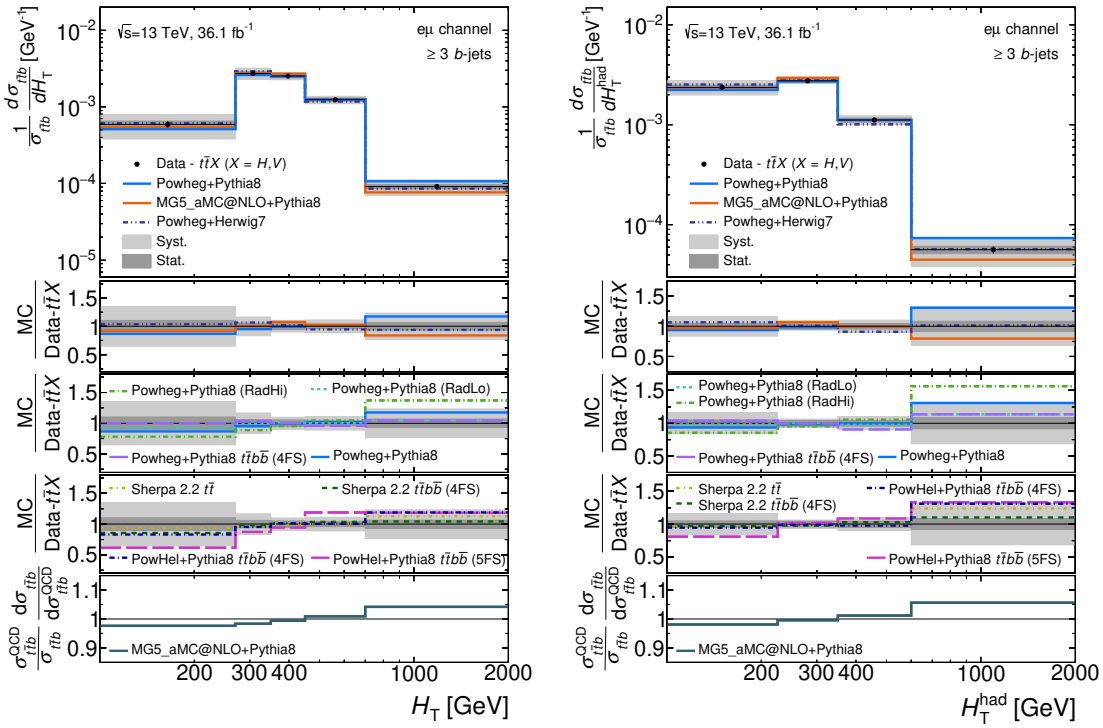


Figure 7.3.: Normalised fiducial cross sections as a function of H_{T}, $H_{\mathrm{T}}^{\mathrm{had}}$ in events with at least 3 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions.

7.2.3. p_{T} of b-jets

The b-jets are arranged in the decreasing order of their p_{T}. To recall, the p_{T} distribution of the highest p_{T} b-jet is symbolised as $p_{\mathrm{T}}^{b_1}$. The b-jets coming from the top-quark decays have a tendency to be harder than the b-jets from additional b-quark production via gluon splitting. Therefore, the first and second b-jets ordered in p_{T} have relatively higher probability to contain b-jets from the top-quark decays. Whereas, the distributions related to the third b-jet contain mainly b-jets from gluon splitting. The p_{T} distributions of leading, sub-leading and third leading b-jets of the p_{T} ordered b-jets are shown in Figure 7.4 in the $\geq$3b-jets fiducial phase space. The quantitative assessment of the data-MC agreement are shown in Table 7.5. Except for the POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS) for the leading and third leading p_{T} b-jets, all the MC predictions describe the data well. The measurement uncertainties are between 10% and 25% depending on the p_{T} of the b-jets.

Generators	$p_{\mathrm{T}}^{b_1}$		$p_{\mathrm{T}}^{b_2}$		$p_{\mathrm{T}}^{b_3}$	
	χ^2/ndof	p-value	χ^2/ndof	p-value	χ^2/ndof	p-value
POWHEG+PYTHIA 8	2.09 / 4	0.72	0.50 / 3	0.92	0.09 / 2	0.95
MG5_aMC@NLO+PYTHIA8	2.62 / 4	0.62	0.27 / 3	0.97	0.33 / 2	0.85
SHERPA2.2 $t\bar{t}$	0.98 / 4	0.91	0.67 / 3	0.88	0.02 / 2	0.99
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	3.52 / 4	0.47	0.68 / 3	0.88	0.21 / 2	0.90
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	10.9 / 4	0.03	2.58 / 3	0.46	3.91 / 2	0.14
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	6.21 / 4	0.18	1.96 / 3	0.58	1.30 / 2	0.52
POWHEG+HERWIG7	1.16 / 4	0.89	1.02 / 3	0.80	0.02 / 2	0.99
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	2.62 / 4	0.62	0.53 / 3	0.91	0.46 / 2	0.80
POWHEG+PYTHIA 8 (RadHi)	2.71 / 4	0.61	0.56 / 3	0.91	0.26 / 2	0.88
POWHEG+PYTHIA 8 (RadLo)	1.93 / 4	0.75	0.64 / 3	0.89	0.05 / 2	0.97

Table 7.5.: Values of χ^2 per degree of freedom and p-values between the unfolded normalised cross-section and the predictions for b-jet p_{T} measurements. The number of degrees of freedom is equal to the number of bins minus one.

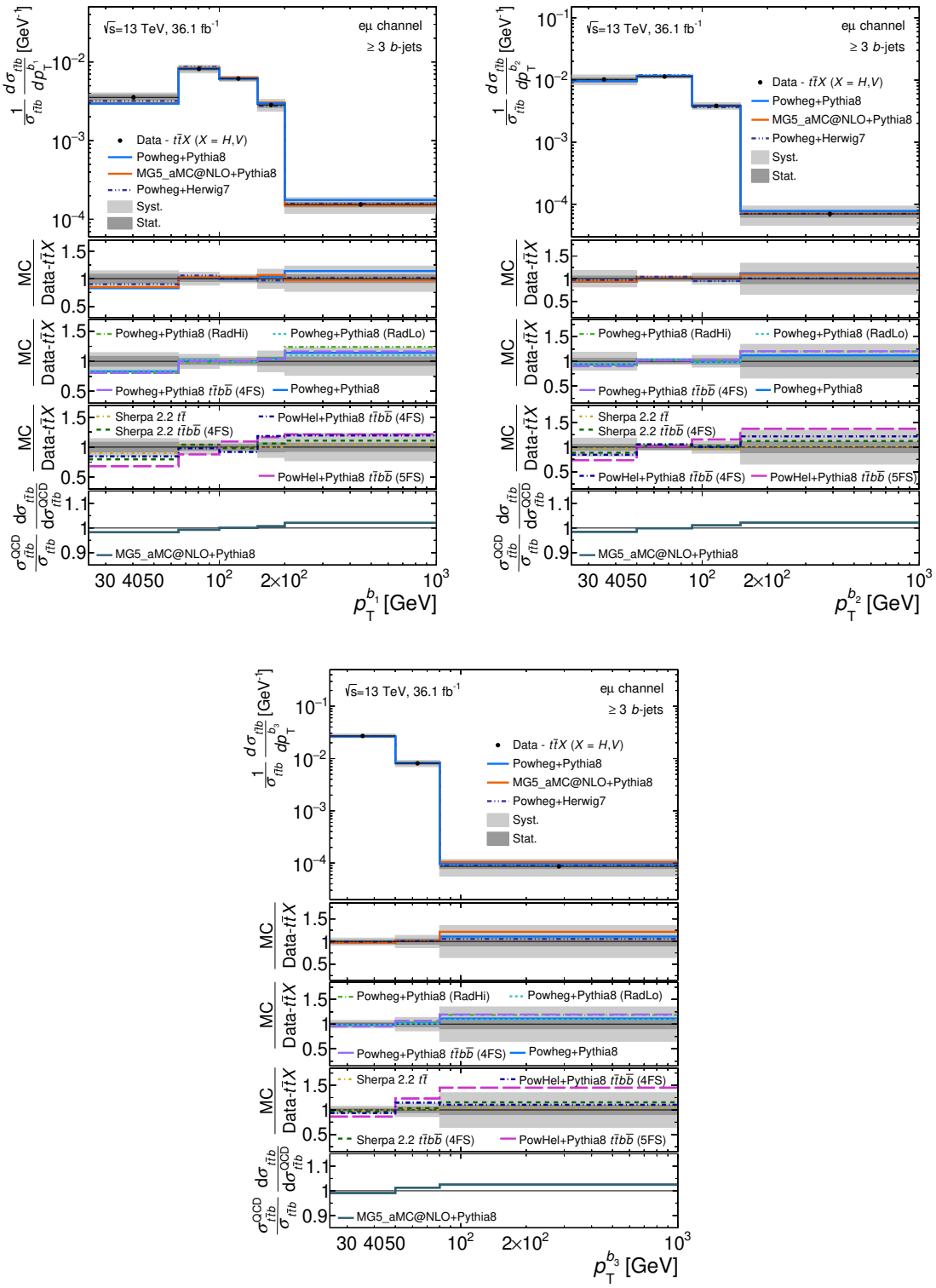


Figure 7.4.: Normalised fiducial cross sections as a function of p_T for p_T ordered jets in events with at least 3 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions.

7.2.4. Properties of Leading *b*-jets System

The b_1b_2 system is built from the two highest p_T b-jets and their properties are studied as a function of the invariant mass, p_T and angular distance as presented in Figure 7.6 for the ≥ 3 b-jets fiducial phase space. The p_T of the b_1b_2 system is measured with a precision of 10%-15% over the full range. It is well described by the different MC predictions, which vary significantly less than the experimental uncertainty. The distributions of ΔR for the two b-jets and the invariant mass of the b_1b_2 pair are measured with slightly higher uncertainties and also show little variation between the different predictions. Good agreement between the data and the models is confirmed by the p-values listed in Table 7.6.

Generators	$m_{bb}^{\Delta\text{min}}$		$p_{T,bb}^{\Delta\text{min}}$		$\Delta R_{bb}^{\Delta\text{min}}$	
	χ^2/ndof	p-value	χ^2/ndof	p-value	χ^2/ndof	p-value
POWHEG+PYTHIA 8	1.55 / 4	0.82	1.74 / 3	0.63	0.70 / 4	0.95
MG5_aMC@NLO+PYTHIA8	1.73 / 4	0.79	1.08 / 3	0.78	3.73 / 4	0.44
SHERPA2.2 $t\bar{t}$	0.25 / 4	0.99	0.64 / 3	0.89	0.99 / 4	0.91
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	2.88 / 4	0.58	0.76 / 3	0.86	2.88 / 4	0.58
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	3.74 / 4	0.44	4.75 / 3	0.19	4.70 / 4	0.32
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	1.35 / 4	0.85	2.90 / 3	0.41	0.86 / 4	0.93
POWHEG+HERWIG7	0.48 / 4	0.98	0.42 / 3	0.94	0.97 / 4	0.91
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	1.89 / 4	0.76	0.79 / 3	0.85	0.68 / 4	0.95
POWHEG+PYTHIA 8 (RadHi)	3.77 / 4	0.44	3.49 / 3	0.32	0.50 / 4	0.97
POWHEG+PYTHIA 8 (RadLo)	1.04 / 4	0.90	0.95 / 3	0.81	1.01 / 4	0.91

Table 7.6.: Values of χ^2 per degree of freedom and p-values between the unfolded normalised cross-section and the predictions for $\Delta R_{bb}^{\Delta\text{min}}$, $p_{T,bb}^{\Delta\text{min}}$ and $m_{bb}^{\Delta\text{min}}$. The number of degrees of freedom is equal to the number of bins minus one.

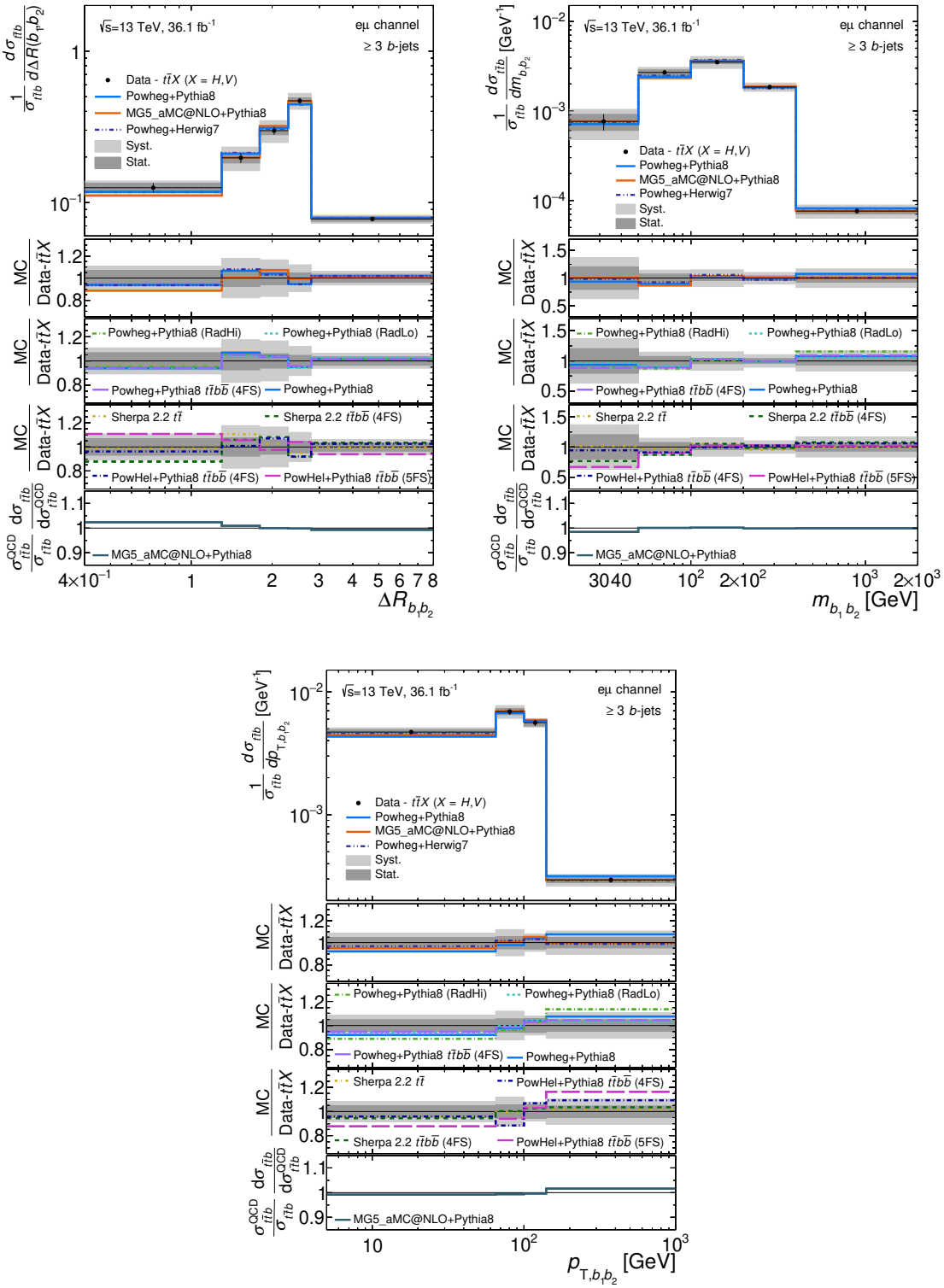


Figure 7.5.: Normalised fiducial cross sections as a function of the angular correlation, $\Delta R_{b_1 b_2}$, the mass, $m_{b_1 b_2}$, and the p_{T} of the system of the two highest p_{T} b-jets for events with at least 3 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions.

7.2.5. Properties of Closest *b*-jets

The same set of observables is investigated for the *b*-jet pair reconstructed from the two closest *b*-jets ($bb_{\Delta\text{min}}$) in the events which implies the *b*-jets having minimum ΔR. The ΔR, invariant mass and p_{T} of the $bb_{\Delta\text{min}}$ system is shown in Figure 7.6 for the ≥ 3 *b*-jets selection. The uncertainties are similar to those using the *b*-jet pair with the highest p_{T}. However, the model variations are larger.

Generators	$m_{bb}^{\Delta\text{min}}$		$p_{\text{T},bb}^{\Delta\text{min}}$		$\Delta\text{R}_{bb}^{\Delta\text{min}}$	
	χ^2/ndof	p-value	χ^2/ndof	p-value	χ^2/ndof	p-value
POWHEG+PYTHIA 8	1.37 / 4	0.85	0.42 / 4	0.98	0.78 / 3	0.86
MG5_aMC@NLO+PYTHIA8	3.67 / 4	0.45	2.50 / 4	0.65	1.22 / 3	0.75
SHERPA2.2 $t\bar{t}$	0.17 / 4	1.0	0.06 / 4	1.0	0.99 / 3	0.80
SHERPA2.2 $t\bar{t}b\bar{b}$ (4FS)	1.36 / 4	0.85	0.52 / 4	0.97	0.21 / 3	0.98
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS)	0.18 / 4	1.0	12.7 / 4	0.01	27.9 / 3	< 0.01
POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (4FS)	4.29 / 4	0.37	2.36 / 4	0.67	0.81 / 3	0.85
POWHEG+HERWIG7	0.87 / 4	0.93	0.06 / 4	1.0	0.95 / 3	0.81
POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS)	1.12 / 4	0.89	1.00 / 4	0.91	0.30 / 3	0.96
POWHEG+PYTHIA 8 (RadHi)	1.94 / 4	0.75	1.31 / 4	0.86	0.51 / 3	0.92
POWHEG+PYTHIA 8 (RadLo)	0.99 / 4	0.91	0.28 / 4	0.99	0.86 / 3	0.84

Table 7.7.: Values of χ^2 per degree of freedom and p-values between the unfolded normalised cross-section and the predictions for $\Delta\text{R}_{bb}^{\Delta\text{min}}$, $p_{\text{T},bb}^{\Delta\text{min}}$ and $m_{bb}^{\Delta\text{min}}$. The number of degrees of freedom is equal to the number of bins minus one.

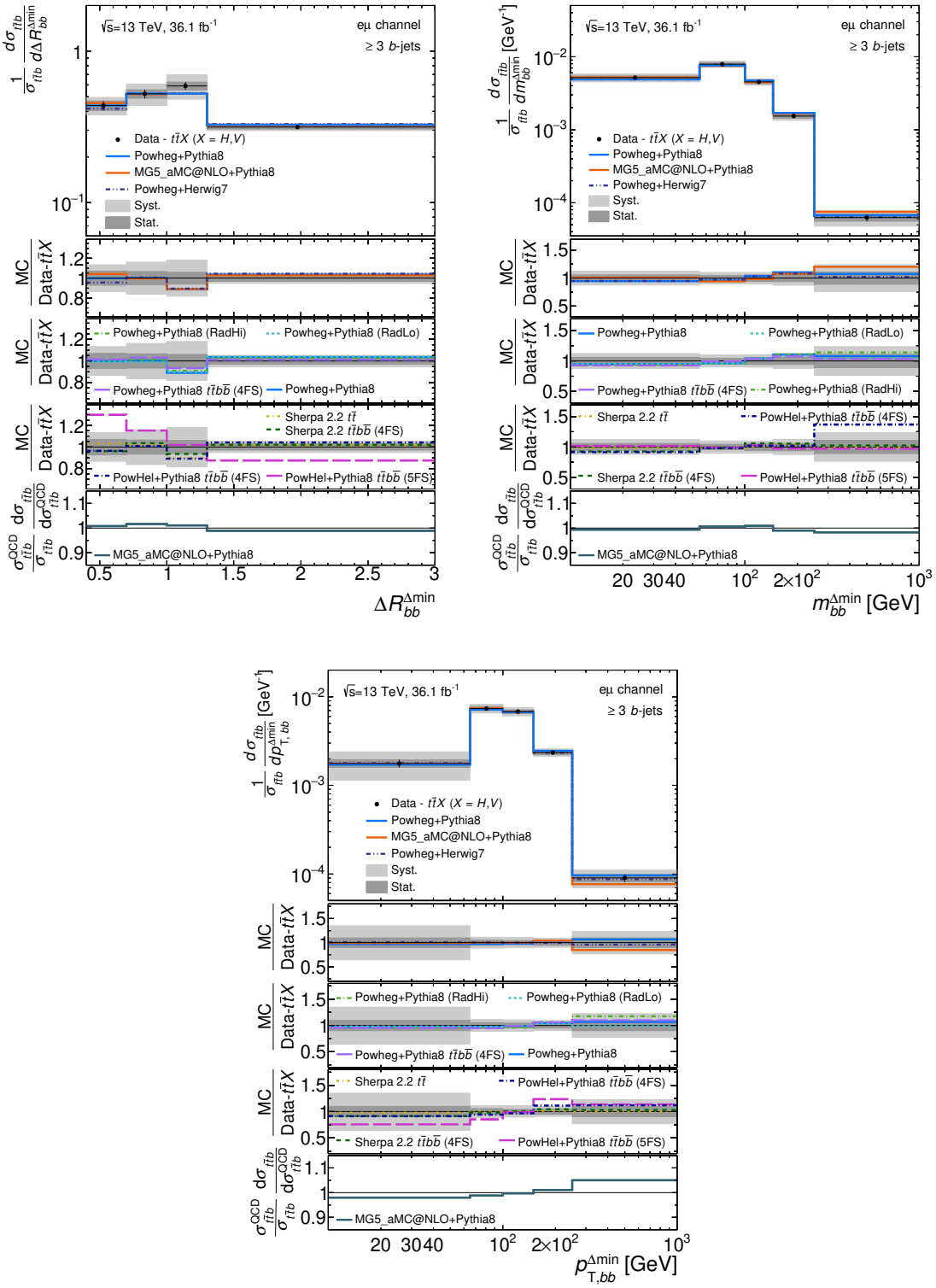


Figure 7.6.: Normalised fiducial cross sections as a function of minimum $\Delta R_{bb}^{\Delta\text{min}}$ and the mass, m_{bb}, and system p_{T} for the closest pair of b-jets in events with at least 3 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions

It is observed that all the MC samples describe the data well within total uncertainties except the POWHEL+PYTHIA8 $t\bar{t}b\bar{b}$ (5FS) in some cases. Both the $t\bar{t}$ and $t\bar{t}b\bar{b}$ predictions describe the measurement equally well. Adding more partons in the matrix element at NLO improves the MC modelling as can be seen from Sherpa 2.2 $t\bar{t}$. The Sherpa 2.2 $t\bar{t}$ and Sherpa 2.2 $t\bar{t}b\bar{b}$ also differ in treating the b-quarks as massless and massive respectively. By comparing these MC models, it can be concluded that adding more partons in matrix element at NLO improves the modelling whereas treating the b-quarks as massive particles is not that important. This would be because of high p_{T} cut on the jets (25 GeV) as compared to the mass of the b-quarks. The same can be deduced by comparing POWHEG+PYTHIA 8 $t\bar{t}$ and POWHEG+PYTHIA 8 $t\bar{t}b\bar{b}$ (4FS).

The breakdown of the uncertainties on all the differential distributions in each bin are presented in the tables in Appendix C. The figures in Section 6.5 and Appendix B show the contribution of major uncertainty sources in the differential results.

As mentioned before, the b-jets in the b_1b_2 system are expected to originate from the top-quark pair whereas the $bb_{\Delta\mathrm{min}}$ system is expected to contain b-jets from gluon splitting which are very collinear and soft. Therefore, the observables involving the two closest b-jets are best to interpret in events with at least 4 b-jets as one would expect this pair of b-jets to originate from gluon splitting. An attempt is made to show the kinematic properties of the b_1b_2 and $bb_{\Delta\mathrm{min}}$ systems for the ≥ 4 b-jets fiducial phase space as shown in Figures 7.7- 7.8, measured with an uncertainty of 25-40%. Significantly smaller number of events in this selection leads to high statistical uncertainties in these distributions as can be seen in the tables in Appendix D.

The ΔR distribution of b_1b_2 system shows that the b-jets coming from $t\bar{t}$ system are well separated as it peaks at higher values of ΔR and the invariant mass peaking above 100 GeV as expected (Figure 7.7). The lower statistics doesn't allow finer binning of the distributions for detailed study. The distributions for the $bb_{\Delta\mathrm{min}}$ system can be compared for the $\geq 3b$ and $\geq 4b$ phase spaces in Figures 7.6 and Figures 7.8 respectively. The ΔR distribution shows that as soon as ≥ 4 b-jets are selected the closely lying b-jets are originating from gluon splitting and are very collinear unlike the ≥ 3 b-jets selection where one b-jet is from the top quark. The invariant mass of $bb_{\Delta\mathrm{min}}$ in the $\geq 4b$ phase space should be exponentially decreasing as the events are dominated by $g \to b\bar{b}$ process. These distributions are extremely important as discriminating parameters for the $g \to b\bar{b}$, $H \to b\bar{b}$ and $Z \to b\bar{b}$ processes for future studies with better statistics.

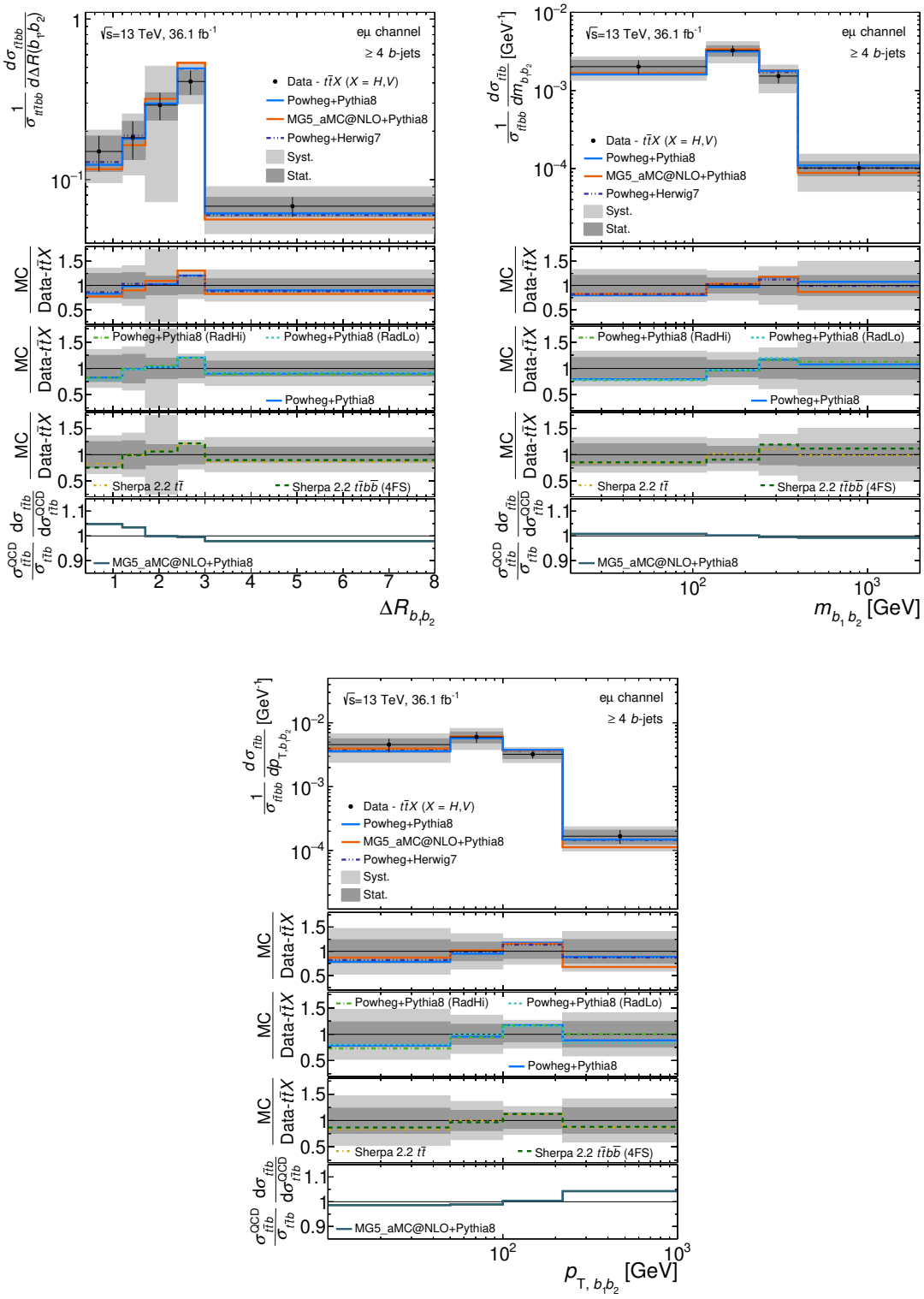


Figure 7.7.: Normalised fiducial cross sections as a function of the angular correlation, $\Delta R_{b_1 b_2}$, the mass, $m_{b_1 b_2}$, and the p_{T} of the system of the two highest p_{T} b-jets in events with at least 4 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions

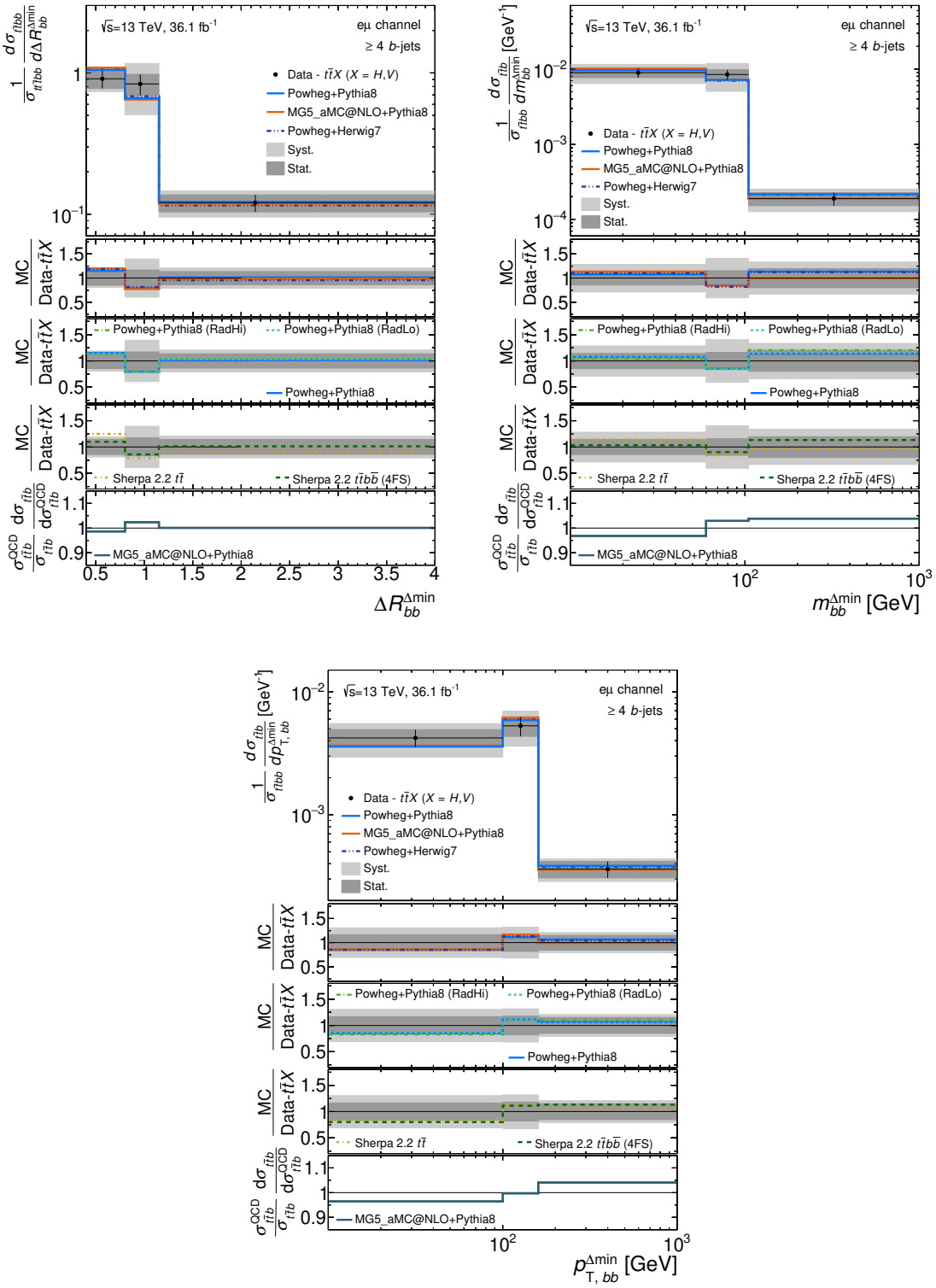


Figure 7.8.: Normalised fiducial cross sections as a function of minimum $\Delta R_{bb}^{\Delta\text{min}}$ and the mass, m_{bb}, and system p_{T} for the closest pair of b-jets in events with at least 4 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}X$ productions

Chapter 8.

Conclusion

This thesis presented the fiducial inclusive and normalised differential cross-sections for the production of a pair of top quarks in association with b-jets at 13 TeV using 36.1 fb^{-1} of pp collision data collected by the ATLAS detector at the LHC. The results are presented for the $e\mu$ channel which contains one electron, one muon and at least two b-jets in the final state. The background from non-$t\bar{t}$ processes giving the same final state signature as the signal are estimated which corresponds to $\sim$4% altogether. The requirement of $t\bar{t}$ with additional b-jets also suffer from the $t\bar{t}c$ and $t\bar{t}l$ background where b-jets are mis-tagged as c- and light-flavoured jets. The data after non-$t\bar{t}$ background subtraction are unfolded to particle level within fiducial phase spaces. The $t\bar{t}c$ and $t\bar{t}l$ background correction factors are also implemented in the unfolding procedure. The fiducial inclusive cross-section obtained for the production of $t\bar{t}b$ and $t\bar{t}b\bar{b}$ final states irrespective of the production mechanism are

$$\begin{aligned} \sigma_{t\bar{t}b} &= 181 \pm 5(\text{stat}) \pm 24(\text{syst}) \text{ fb} \\ \sigma_{t\bar{t}b\bar{b}} &= 27 \pm 3(\text{stat}) \pm 7(\text{syst}) \text{ fb} \end{aligned}$$

The contribution coming only from the QCD process is also estimated by subtracting MC based predictions of $t\bar{t}X$ ($X = H, Z, W$) contributions which results

$$\begin{aligned} \sigma_{t\bar{t}b}^{\text{Data}-t\bar{t}X} &= 177 \pm 5(\text{stat}) \pm 24(\text{syst}) \text{ fb} \\ \sigma_{t\bar{t}b\bar{b}}^{\text{Data}-t\bar{t}X} &= 25 \pm 3(\text{stat}) \pm 7(\text{syst}) \text{ fb} \end{aligned}$$

The cross-section for ≥ 3 b-jets phase space has a smaller uncertainty of 13% as compared to the ≥ 4 b-jets which has 28% uncertainty. The results obtained after subtracting the $t\bar{t}X$ contributions are compared with various $t\bar{t}b\bar{b}$ predictions. The measured $t\bar{t}b$ cross-section is found to be higher than the prediction but lies within 2σ. In case of $t\bar{t}b\bar{b}$, the measurements are slightly higher than the predicted values but compatible within the uncertainties. It can be inferred that the theory describes the data well when both the b-quarks are inside the acceptance. The results obtained by the ATLAS collaboration at 8 TeV [13] for the $t\bar{t}b$ fiducial phase space in

the $e\mu$ channel suffered from 19% statistical uncertainty and in total 32% uncertainty. The $t\bar{t}b\bar{b}$ fiducial cross-section measurement was performed for the full di-lepton decay channel and resulted in 36% uncertainty [13]. The increased centre of mass energy and luminosity as well as improvements in the modelling lead to the lower uncertainties in the presented analysis. It also provides the scope to exploit the differential cross-section to analyse the results in great details.

The differential cross-section as a function of b-jet multiplicity, b-jet p_{T}, kinematic properties of b-jet pairs and global event properties has also been presented for the $t\bar{t}b$ and $t\bar{t}b\bar{b}$ fiducial phase spaces. Due to the lower statistics in the $t\bar{t}b\bar{b}$ case, the results cannot be well interpreted but the results of $t\bar{t}b$ are explained in the thesis. The observables are well described by most of the MC predictions. In general, the differential distributions are measured within a precision of 10% and rise up to 30% at the edges of the phase space for some of them. All the MC models predict too few additional b-jets whereas the b-jets coming from the top quark are well described. Only the Sherpa 2.2 $t\bar{t}$ sample which includes additional jets in the ME calculation describes the data very well for additional b-jet multiplicity and also yields best agreement with data for most of the observables in ≥ 3 b-jets phase space. POWHEL +PYTHIA8 $t\bar{t}b\bar{b}$ (5FS) shows poor agreement for some of the observables such as $p_{\mathrm{T}}^{b_3}$, p_{T,b_1b_2} and $\Delta R_{bb}^{\Delta\mathrm{min}}$. The differential distributions are equally well described by the predictions having $t\bar{t}$ in matrix element calculation and additional b-jets coming from the parton shower and the predictions having full $t\bar{t}b\bar{b}$ in the matrix element calculation. The differential analysis of $t\bar{t}b\bar{b}$ with higher statistics is of great importance and will be extremely interesting to look at, especially the properties of the b-jets and b-jet pairs as presented in this analysis.

It would also be interesting to study both the $t\bar{t}b\bar{b}$ and $t\bar{t}jj$ processes (j represents jets of any flavour) to understand the production of additional jets overall and improve the modelling [173]. Studying the ratio of $t\bar{t}b\bar{b}/t\bar{t}jj$ reduces several experimental uncertainties by cancellation such as (JES/JER). The analysis presented in this thesis is systematics dominated but gaining more statistics in data and MC would help to improve the systematic uncertainties coming from modelling as well, therefore, improving the results overall. MC generators which keep improving the calculations require these experimental inputs for further betterment of the predictions.

Appendix A.

Unfolding Checks

The remaining results of the checks performed for validation of unfolding procedure in Section 6.3 is presented here.

A.1. Pull-test and Optimisation of binning

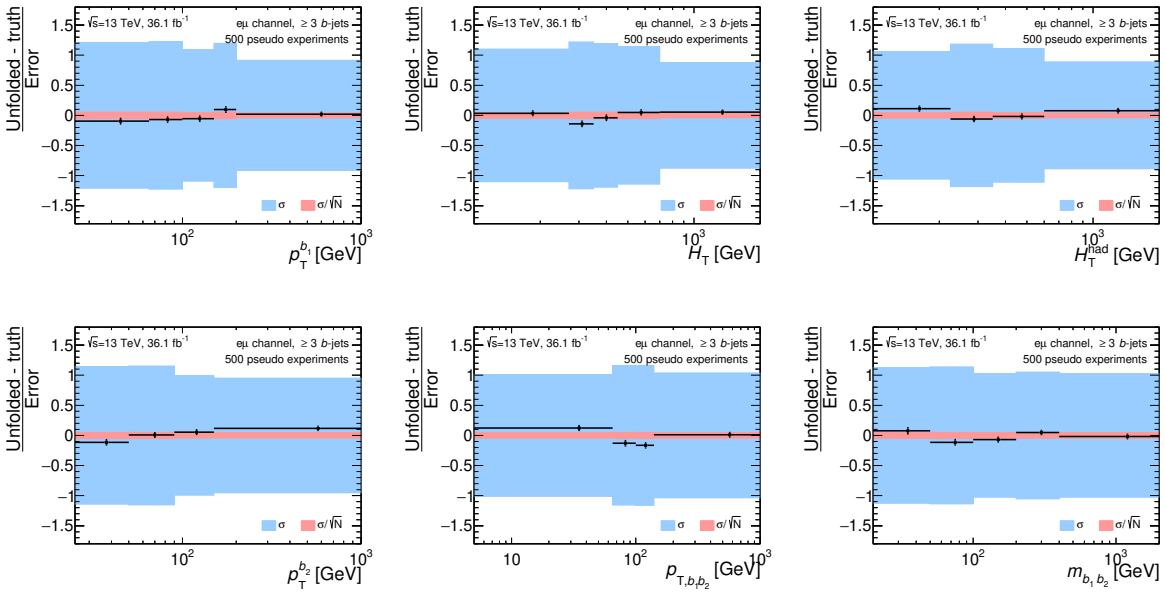


Figure A.1.: The results of the pull tests performed with pseudo-data for each observable $N_{b\text{-jets}}$, $p_T^{b_1}$, $p_T^{b_2}$, $p_T^{b_3}$, H_T, H_T^{had}, $\Delta R_{b_1b_2}$, p_{T,b_1b_2}, $m_{b_1b_2}$

A.2. Optimisation of number of iterations

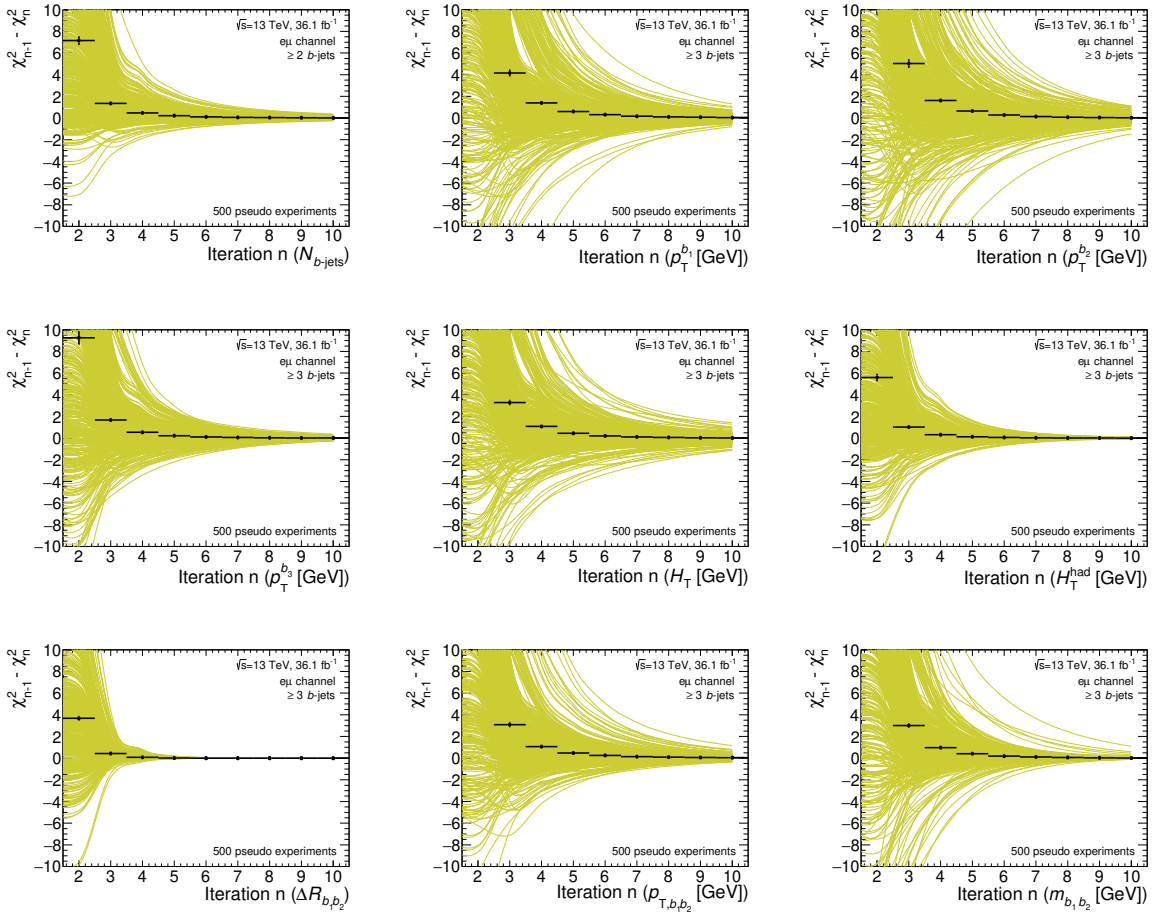


Figure A.2.: Change in χ^2 after n^{th} iteration of Bayesian unfolding in various pseudo-experiments shown in yellow, where the χ^2 test is performed between unfolded and generated particle-level distribution taking into account the covariance from unfolding. The black lines shows the mean change in χ^2 and results are presented for: $N_{b\text{-jets}}$, $p_T^{b_1}$, $p_T^{b_2}$, $p_T^{b_3}$, H_T, H_T^{had}, $\Delta R_{b_1 b_2}$, $p_{T,b_1 b_2}$ and $m_{b_1 b_2}$.

A.3. Stress Test

Results are showing the bias distribution for the stress test for various observables. In each of the plots the input distributions of POWHEG+PYTHIA 8 are re-weighted to POWHEG+PYTHIA 8 (RadHi) (left) to POWHEG+HERWIG7 (middle) and MG5_aMC@NLO+PYTHIA8 (right). The red points show the relative difference in shapes of generated particle-level distributions between POWHEG+PYTHIA 8 and alternative MC, while the black points show the relative bias after unfolding.

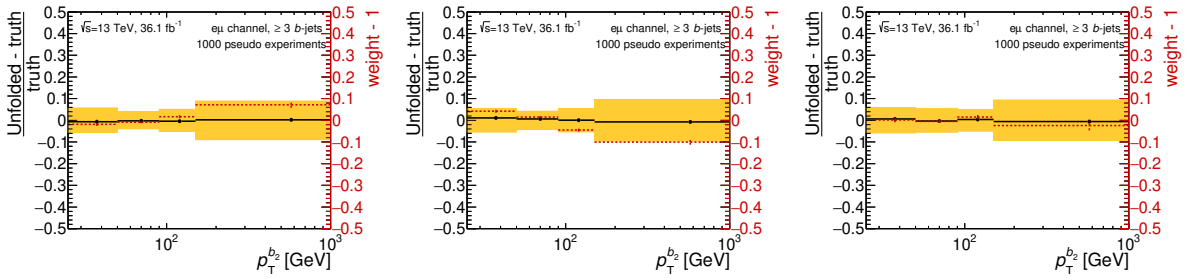


Figure A.3.: Results showing the bias distribution for the stress test for second leading b-jet p_T spectra.

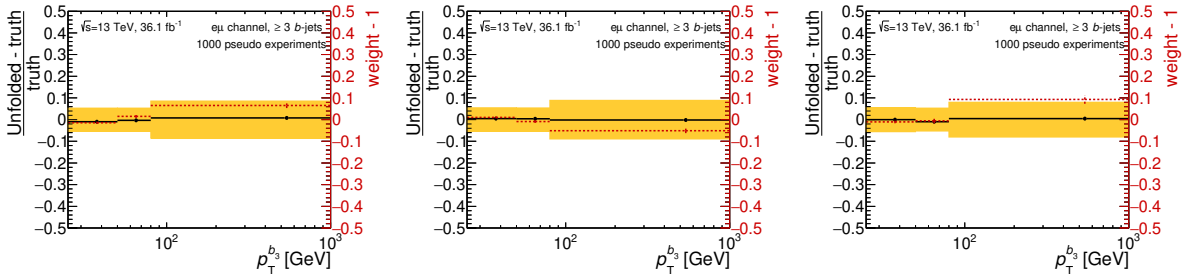


Figure A.4.: Results showing the bias distribution for the stress test for third leading b-jet p_T spectra.

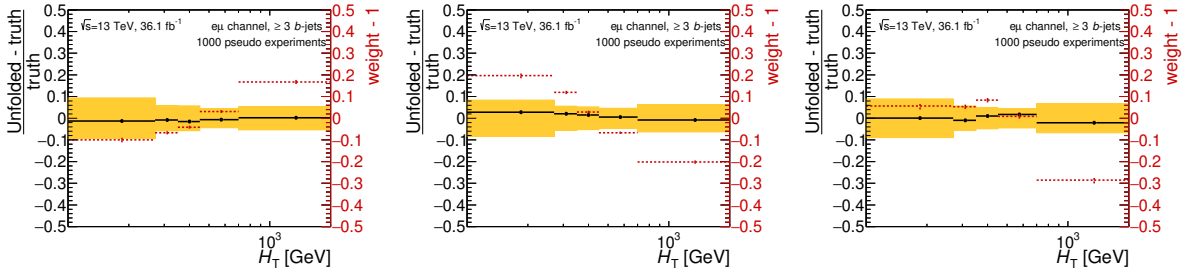


Figure A.5.: Results showing the bias distribution for the stress test for H_T.

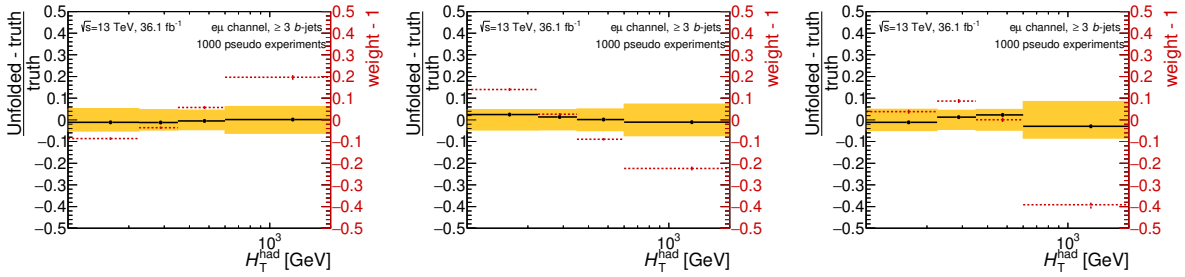


Figure A.6.: Results showing the bias distribution for the stress test for $H_{\mathrm{T}}^{\mathrm{had}}$.

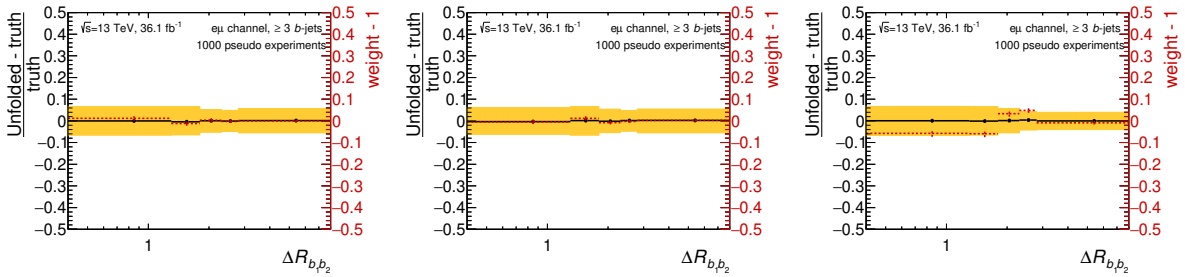


Figure A.7.: Results showing the bias distribution for the stress test for $\Delta \mathrm{R}_{b_1 b_2}$ spectra.

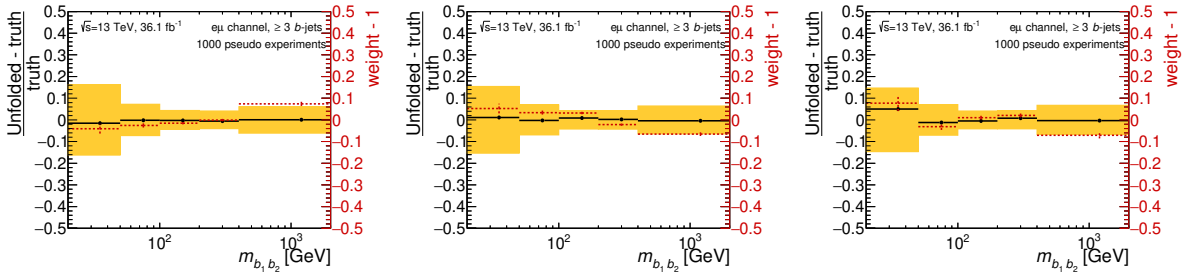


Figure A.8.: Results showing the bias distribution for the stress test for $m_{b_1 b_2}$.

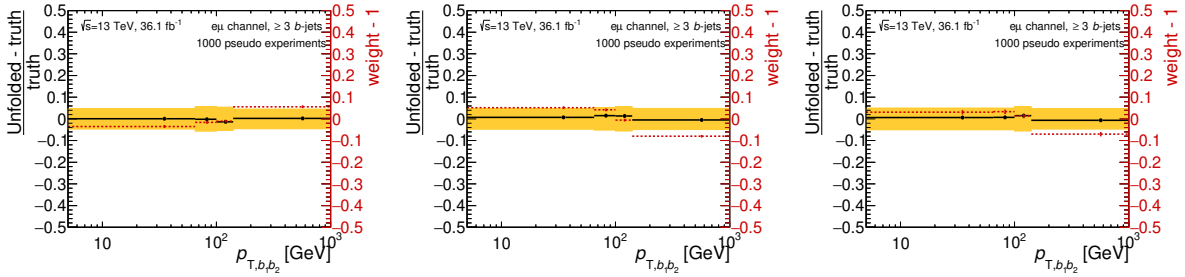


Figure A.9.: Results showing the bias distribution for the stress test for $p_{\mathrm{T},b_1 b_2}$.

Appendix B.

Statistical and Systematic Uncertainties

B.1. Components of Systematic Uncertainties

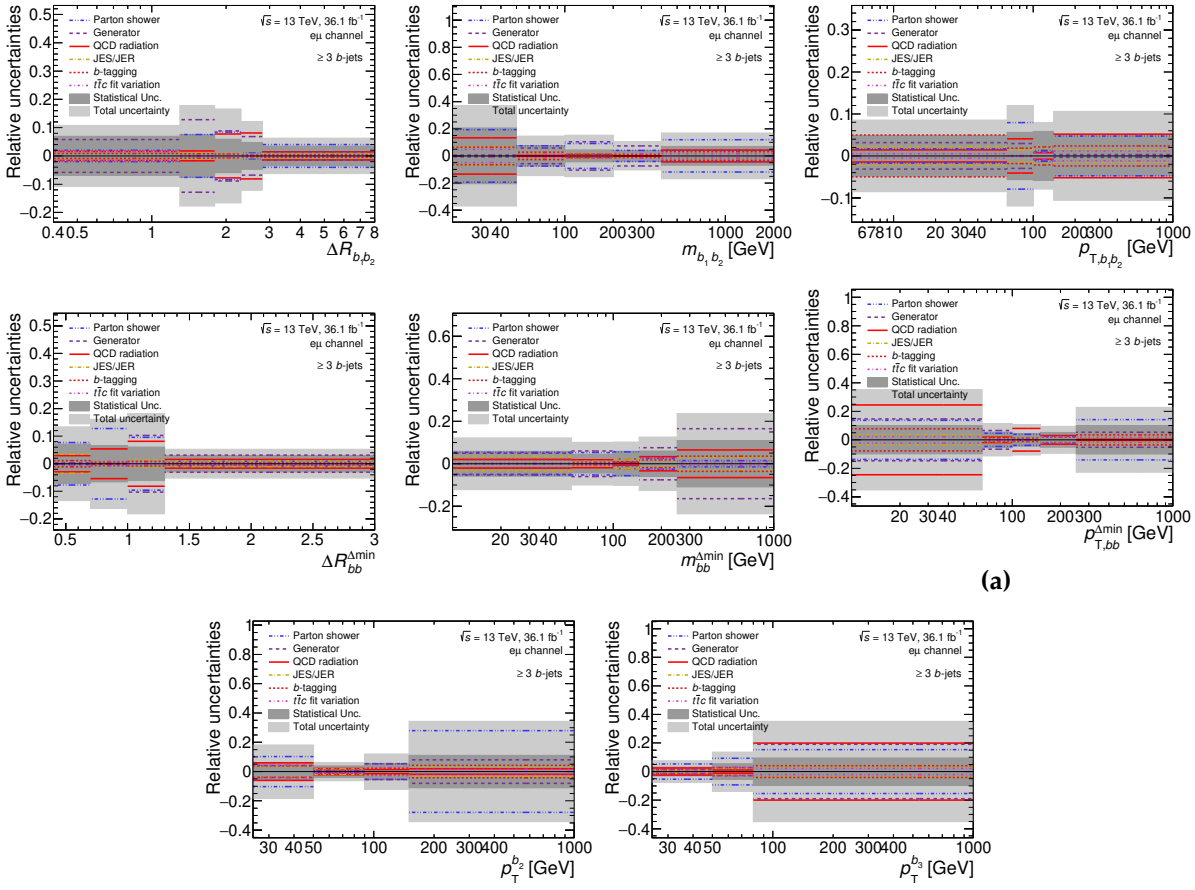


Figure B.1.: Relative systematic uncertainties from various theoretical and experimental sources for different observables.

B.2. Correlation Matrices for Statistical and Systematic Uncertainties

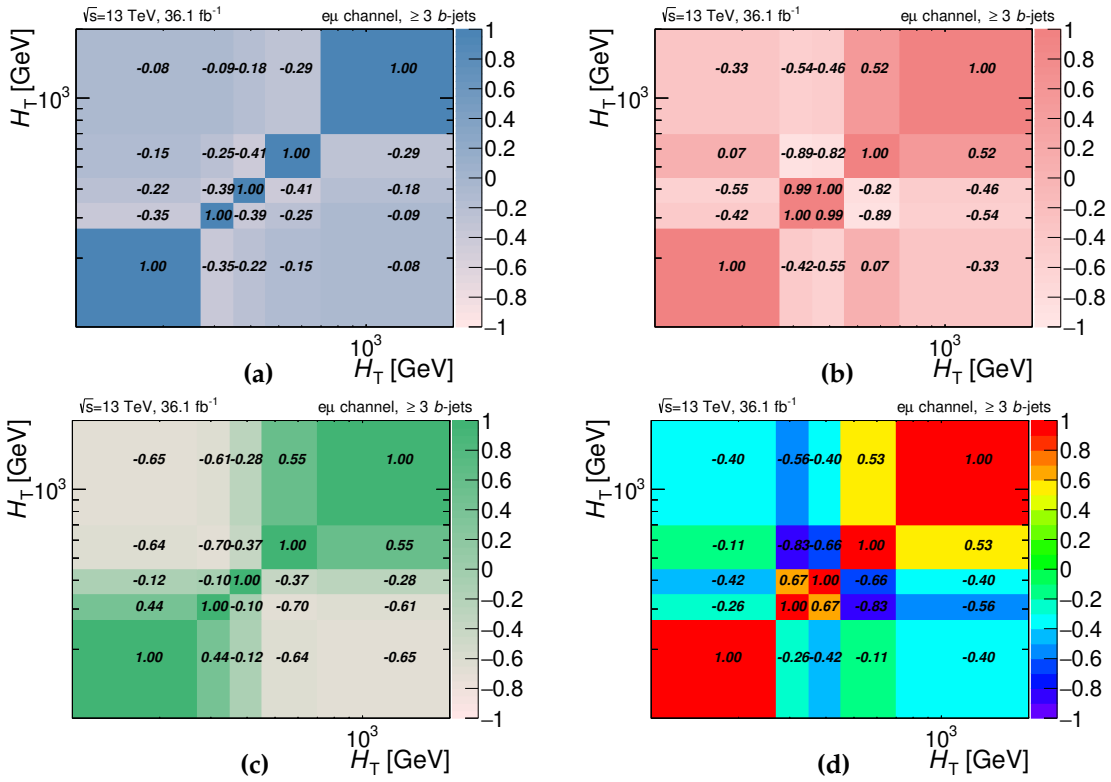


Figure B.2.: Correlation between bins of H_{T} normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

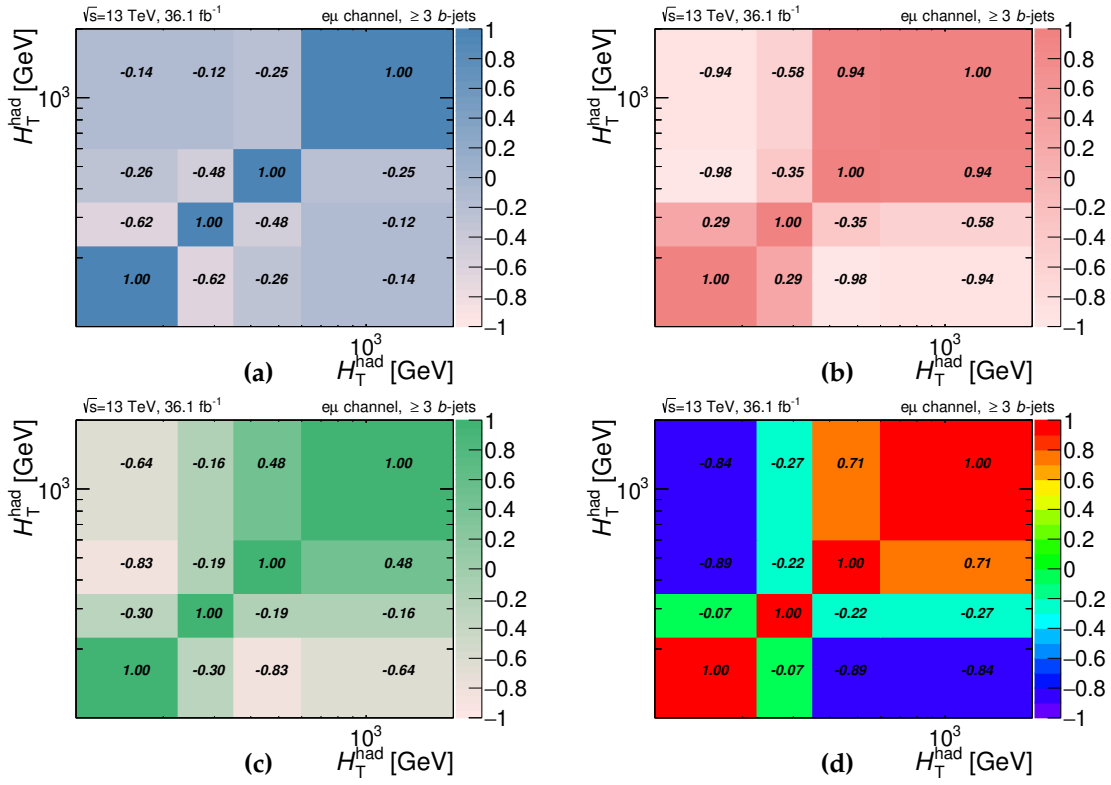


Figure B.3.: Correlation between bins of $H_{\mathrm{T}}^{\mathrm{had}}$ normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics

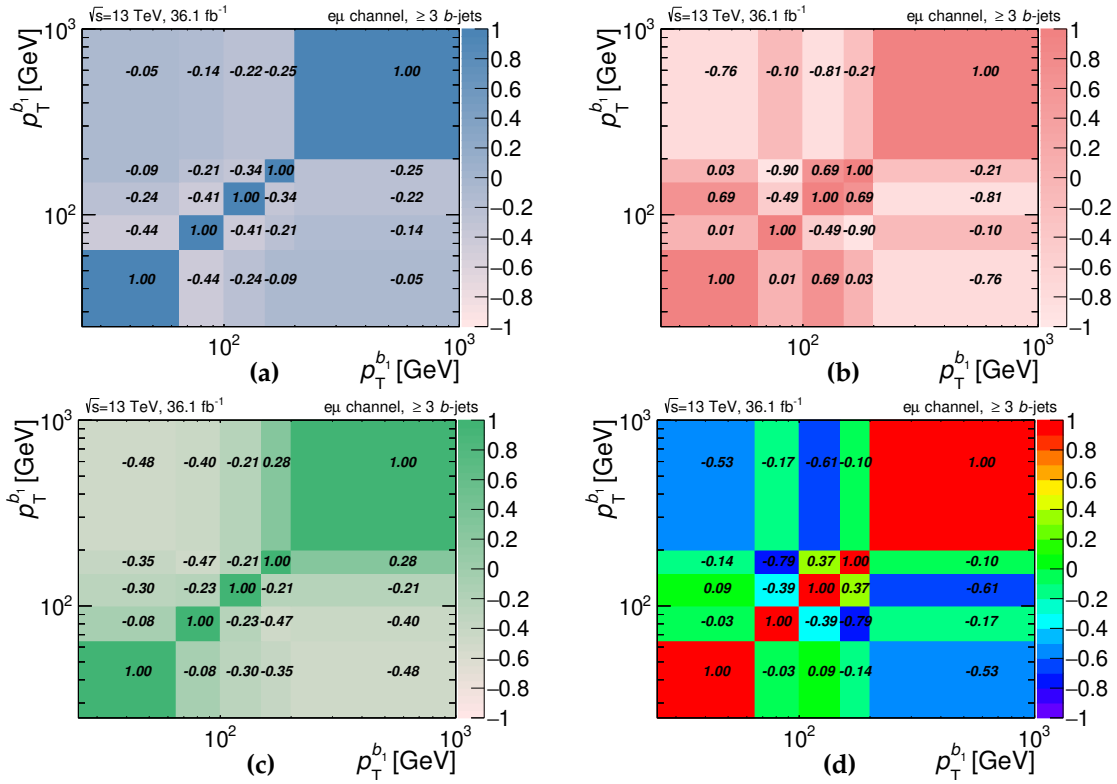


Figure B.4.: Correlation between bins of leading b-jet p_T normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

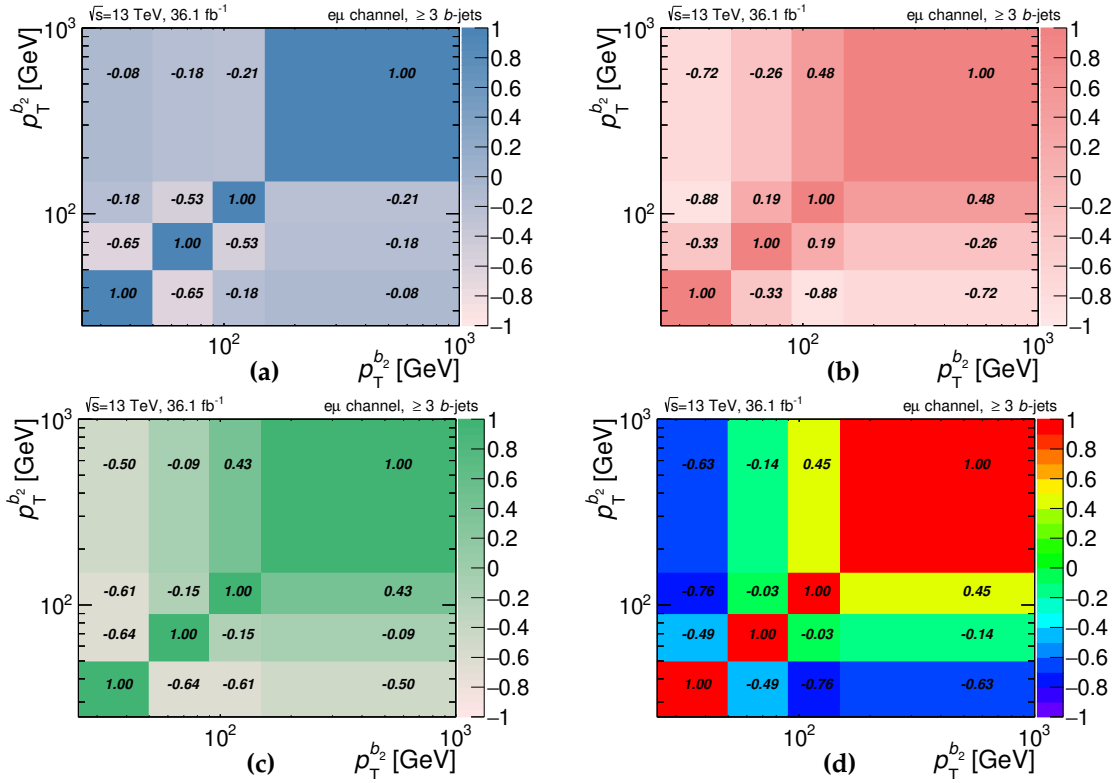


Figure B.5.: Correlation between bins of second leading b-jets p_T normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

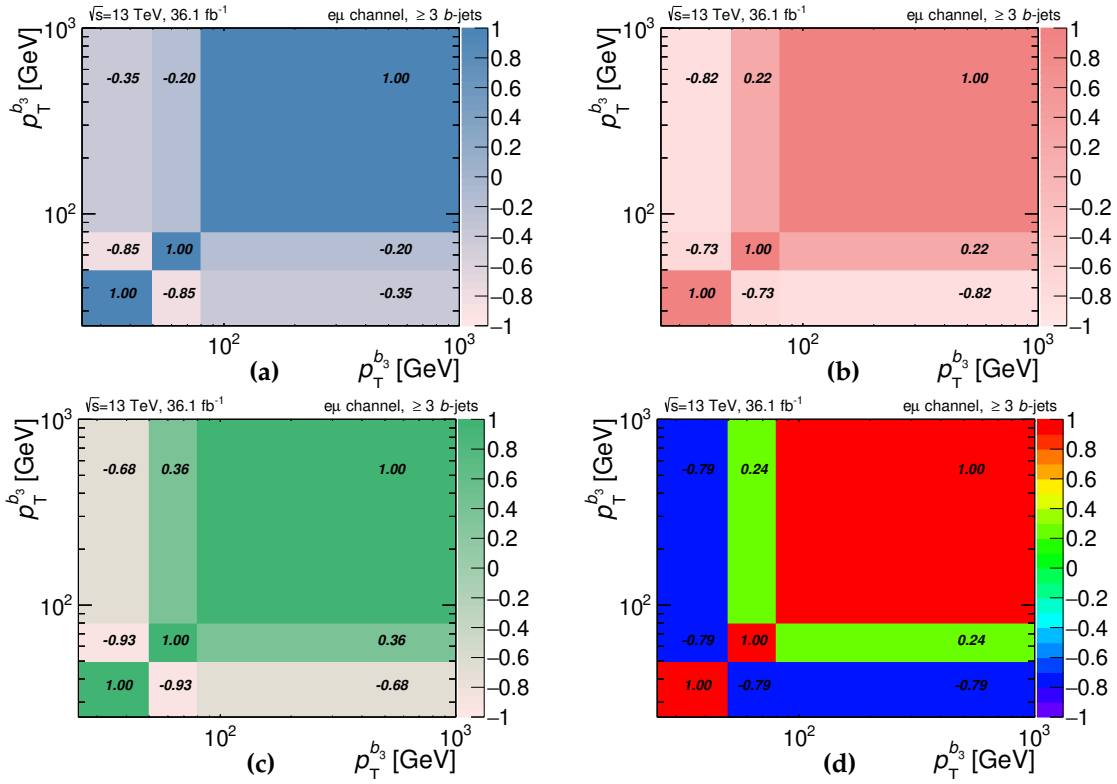


Figure B.6.: Correlation between bins of third leading b-jets p_{T} normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

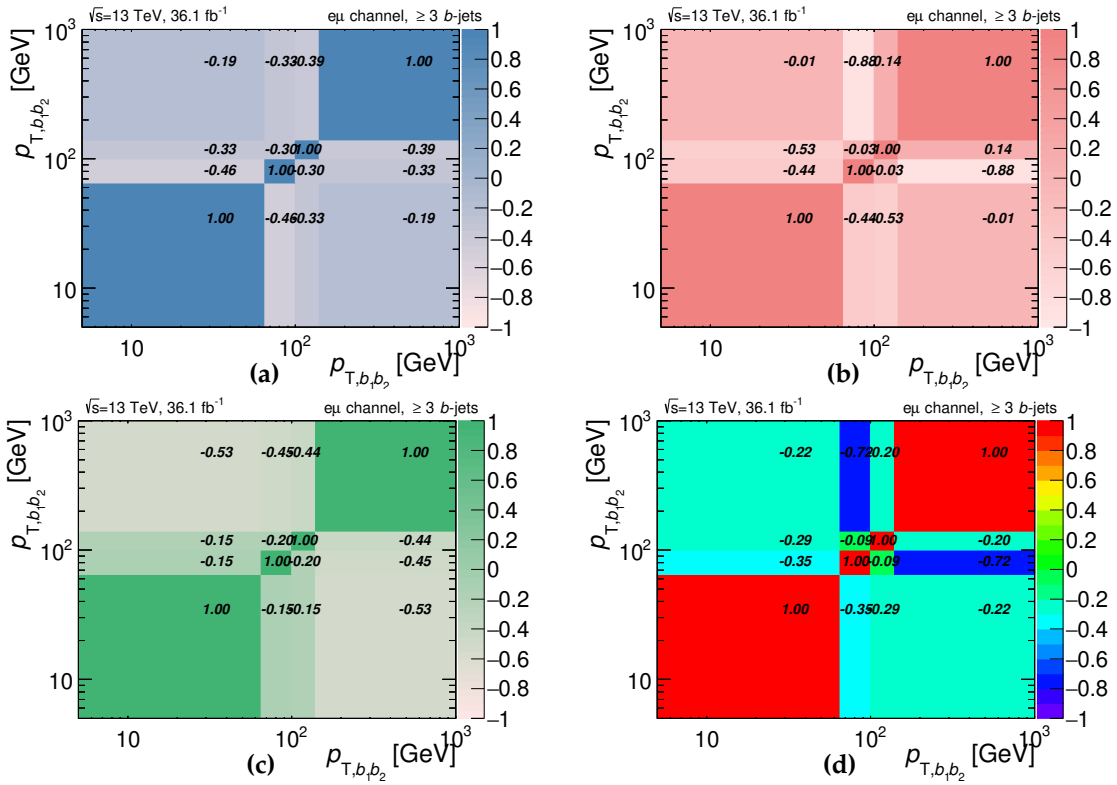


Figure B.7.: Correlation between bins of p_{T,b_1b_2} normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

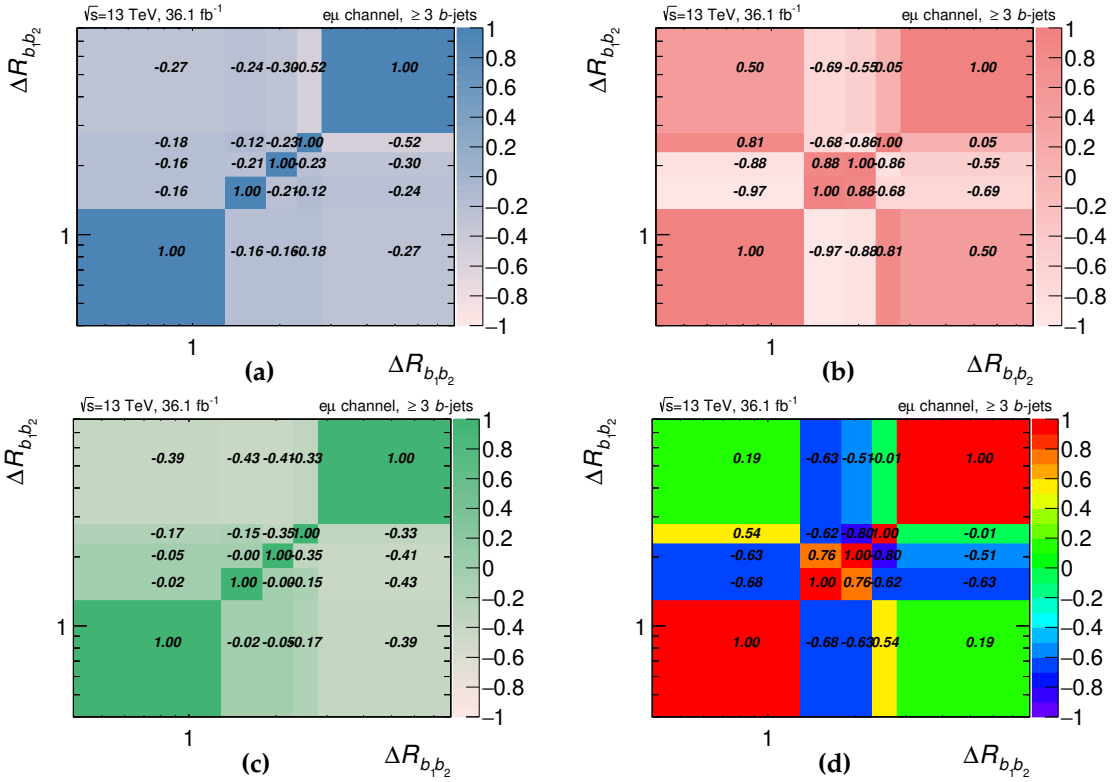


Figure B.8.: Correlation between bins of $\Delta R_{b_1 b_2}$ normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

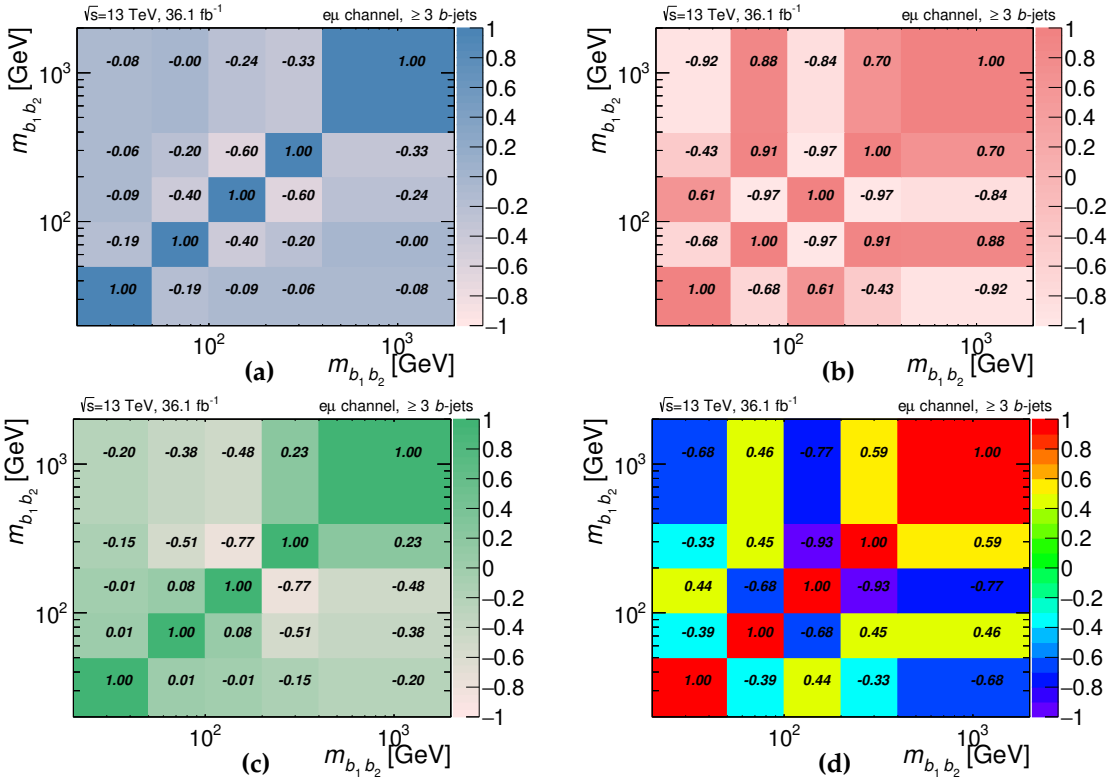


Figure B.9.: Correlation between bins of $m_{b_1 b_2}$ normalised distribution. (a) Statistical only, (b) modeling systematics, (c) detector systematics, (d) Statistical+Systematics.

Appendix C.

Tables of Systematic Uncertainties

Tables C.1- C.11 show the breakdown of systematic uncertainties of normalised cross section measurements in ≥ 3 b-jets fiducial phase space.

$p_{\mathrm{T}}^{b_1}$ [GeV]	(25-65)	(65-100)	(100-150)	(150-200)	(200-1000)
Measured Value [GeV^{-1}]	0.0035	0.0082	0.0061	0.0029	0.0002
Powheg+Pythia8 [GeV^{-1}]	0.0029	0.0082	0.0061	0.0030	0.0002
Uncertainty	%	%	%	%	%
Data statistics	8.26	5.01	4.60	7.05	6.69
MC stats.	4.39	5.20	5.22	5.80	10.55
c-jet shape variation	-2.54	-1.40	0.21	2.09	3.10
JES	3.20	1.76	1.04	2.39	2.95
JER	3.70	0.95	0.58	0.29	2.11
b-tagging calibration	4.32	1.35	0.85	1.88	3.35
lepton (reco+ID+iso.)	0.18	0.11	0.08	0.20	0.19
lepton trigger	0.01	0.00	0.00	0.01	0.01
JVT	0.07	0.01	0.01	0.02	0.03
Pileup reweighting	3.27	0.93	0.98	2.24	2.07
non-$t\bar{t}$ bkg	2.07	0.48	0.81	1.18	3.80
Tot. detector syst.	10.10	4.90	3.70	7.90	9.90
shower	-2.12	-1.47	-2.99	-3.52	16.84
generator	-4.70	-4.51	-0.39	7.81	7.32
$t\bar{t}\,V$	0.00	-0.00	0.02	-0.02	-0.01
$t\bar{t}\,H$	0.24	0.16	-0.06	-0.22	-0.16
ISR/FSR	-3.17	7.66	-3.57	-10.13	5.37
PDF	0.73	0.54	0.76	0.71	2.61
Tot. modeling syst.	6.10	9.03	4.73	13.29	19.31
Total syst.	11.80	10.27	6.01	15.46	21.70
Total	14.40	11.43	7.57	16.99	22.71

Table C.1.: Systematic uncertainties on leading b-jet normalised differential cross section $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dp_{\mathrm{T}}^{b_1}}\right)$ in ≥ 3 b-jet events.

$p_{\mathrm{T}}^{b_2}$ [GeV]	(25 - 50)	(50 - 90)	(90 - 150)	(150-1000)
Measured Value [GeV^{-1}]	0.0101	0.0114	0.0039	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0095	0.0116	0.0038	0.0001
Uncertainty	%	%	%	%
Data statistics	6.08	3.75	5.36	10.63
MC stats.	5.34	5.94	8.54	4.61
c-jet shape variation	-5.46	0.75	3.54	3.50
JES	3.50	1.89	2.48	3.53
JER	4.99	1.74	0.56	3.78
b-tagging calibration	3.89	1.03	2.20	4.22
lepton (reco+ID+iso.)	0.06	0.09	0.13	0.15
lepton trigger	0.01	0.00	0.00	0.01
JVT	0.08	0.02	0.04	0.04
Pileup reweighting	1.50	1.25	1.30	5.74
non-$t\bar{t}$ bkg	0.77	0.21	0.87	2.52
Tot. detector syst.	11.30	4.40	7.30	14.20
shower	-10.30	-0.59	5.33	27.91
generator	-4.08	0.62	5.44	-8.04
$t\bar{t}\,V$	-0.02	0.00	-0.02	0.11
$t\bar{t}\,H$	0.22	0.04	-0.21	-0.31
ISR/FSR	-6.05	2.08	1.39	1.86
PDF	1.96	0.25	1.16	1.39
Tot. modeling syst.	12.78	2.26	7.83	29.14
Total syst.	17.06	4.95	10.70	32.42
Total	18.11	6.21	11.97	34.12

Table C.2.: Systematic uncertainties on subleading b-jet normalised differential cross section $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dp_{\mathrm{T}}^{b_2}}\right)$ in events with ≥ 3 b-jets.

$p_{\mathrm{T}}^{b_3}$ [GeV]	(25 - 50)	(50 - 80)	(80 - 1000)
Measured Value [GeV^{-1}]	0.0269	0.0082	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0265	0.0083	0.0001
Uncertainty	%	%	%
Data statistics	3.02	5.32	9.10
MC stats.	5.49	5.81	4.32
c-jet shape variation	-1.99	3.80	4.91
JES	1.15	2.32	3.46
JER	0.32	0.44	1.19
b-tagging calibration	1.51	2.71	4.13
lepton (reco+ID+iso.)	0.06	0.17	0.25
lepton trigger	0.00	0.01	0.01
JVT	0.03	0.05	0.07
Pileup reweighting	0.81	2.52	3.51
non-$t\bar{t}$ bkg	0.81	1.32	2.78
Tot. detector syst.	3.50	7.80	11.80
shower	-5.34	9.31	15.36
generator	-1.34	-2.81	19.04
$t\bar{t}\,V$	0.01	-0.01	-0.05
$t\bar{t}\,H$	0.20	-0.31	-0.53
ISR/FSR	-2.46	-1.07	19.97
PDF	0.38	0.26	2.20
Tot. modeling syst.	6.04	9.79	31.66
Total syst.	6.98	12.52	33.79
Total	7.61	13.60	34.99

Table C.3.: Systematic uncertainties of 3rd b-jet normalised differential cross section $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dp_{\mathrm{T}}^{b_3}}\right)$ in events with ≥ 3 b-jets.

H_{T} [GeV]	(100 - 270)	(270 - 350)	(350 - 450)	(450 - 700)	(700 - 2000)
Measured Value [GeV^{-1}]	0.0006	0.0027	0.0025	0.0012	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0005	0.0026	0.0025	0.0013	0.0001
Uncertainty	%	%	%	%	%
Data statistics	10.15	5.83	5.22	4.11	6.37
MC stats	4.27	5.21	5.69	8.87	5.92
c-jet shape variation	-5.66	-2.16	-0.48	2.68	2.57
JES	6.16	2.36	1.13	1.42	4.15
JER	0.52	0.64	1.98	1.31	1.88
b-tagging calibration	3.73	1.43	0.52	0.92	2.46
lepton (reco+ID+iso.)	0.33	0.53	0.25	0.21	0.44
lepton trigger	0.03	0.01	0.00	0.00	0.02
JVT	0.12	0.04	0.00	0.03	0.07
Pileup reweighting	4.23	1.50	1.74	0.91	2.44
non-$t\bar{t}$ bkg	1.15	0.90	0.46	0.51	2.77
Tot. detector syst.	15.70	7.20	5.20	5.60	10.50
shower	-10.87	-6.50	-2.60	3.87	19.69
generator	7.68	-10.35	-4.75	7.80	1.72
$t\bar{t}\,V$	-0.00	-0.05	0.06	-0.03	0.05
$t\bar{t}\,H$	0.40	0.24	-0.05	-0.16	-0.13
ISR/FSR	26.69	-6.14	-3.78	-0.03	0.13
PDF	0.09	0.63	0.14	0.04	0.74
Tot. modeling syst.	29.83	13.70	6.61	8.71	19.78
Total syst.	33.71	15.48	8.41	10.36	22.39
Total	35.20	16.54	9.90	11.14	23.28

Table C.4.: Systematic uncertainties for normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dH_{\mathrm{T}}}\right)$ in events with ≥ 3 b-jets.

$H_{\mathrm{T}}^{\mathrm{had}}$ [GeV]	(100 - 225)	(225 - 350)	(350 - 600)	(600 - 2000)
Measured Value [GeV^{-1}]	0.0023	0.0028	0.0011	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0022	0.0027	0.0011	0.0001
Systematics	%	%	%	%
Data statistics	5.04	4.35	4.43	7.93
MC stats	5.71	5.21	9.02	5.73
c-jet shape variation	-4.22	0.63	2.68	3.20
JES	2.67	1.37	2.80	2.08
JER	1.39	0.30	0.83	2.48
b-tagging calibration	2.54	0.74	1.54	2.57
lepton (reco+ID+iso.)	0.37	0.20	0.13	0.36
lepton trigger	0.19	0.07	0.07	0.08
JVT	0.17	0.04	0.07	0.11
Pileup reweighting	2.14	1.56	1.41	2.31
non-$t\bar{t}$ bkg	0.67	0.37	0.50	3.45
Tot. detector syst.	9.90	4.00	6.40	12.30
shower	-9.44	-1.25	5.27	26.16
generator	-6.87	0.93	3.11	8.22
$t\bar{t}$ V	0.08	-0.03	-0.03	-0.02
$t\bar{t}$ H	0.33	-0.08	-0.17	-0.13
ISR/FSR	-0.80	-0.30	-0.70	4.12
PDF	0.12	0.41	0.25	0.54
Tot. modeling syst.	11.71	1.64	6.17	27.73
Total syst.	15.33	4.32	8.89	30.34
Total	16.14	6.13	9.93	31.36

Table C.5.: Systematic uncertainties for normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dH_{\mathrm{T}}^{\mathrm{had}}}\right)$ in events with ≥ 3 b-jets.

$\Delta R_{b_1 b_2}$	0.4 - 1.3	1.3 - 1.8	1.8 - 2.3	2.3 - 2.8	2.8 - 8.0
Measured Value	0.1270	0.1987	0.2990	0.4691	0.0774
Powheg+Pythia8	0.1179	0.2105	0.3103	0.4449	0.0790
Systemtics	%	%	%	%	%
Data statistics	6.68	7.35	5.79	4.65	3.43
MC stats.	4.09	4.15	5.66	5.35	7.99
c-jet shape variation	-0.36	0.99	1.03	0.18	-0.63
JES	1.53	1.08	1.14	0.60	0.79
JER	0.35	0.86	1.45	0.61	0.52
b-tagging calibration	0.88	0.79	0.69	0.34	0.50
lepton (reco+ID+iso.)	0.13	0.11	0.28	0.11	0.03
lepton trigger	0.01	0.01	0.01	0.00	0.00
JVT	0.01	0.00	0.01	0.00	0.00
Pileup reweighting	1.14	1.01	1.79	1.13	0.70
non-$t\bar{t}$ bkg	0.92	2.39	0.89	0.78	0.73
Tot. detector syst.	5.30	5.70	5.60	3.40	2.60
shower	2.00	-7.58	-8.40	1.05	4.01
generator	5.81	-12.81	-8.85	6.81	1.45
$t\bar{t}\,V$	-0.04	0.01	-0.00	0.01	0.01
$t\bar{t}\,H$	-0.24	-0.25	-0.10	-0.02	0.18
ISR/FSR	1.58	-1.73	-7.75	8.13	-1.43
PDF	0.07	0.40	0.29	0.13	0.08
Tot. modeling syst.	6.35	14.99	14.45	10.66	4.51
Total syst.	8.27	16.04	15.50	11.19	5.20
Total	10.63	17.64	16.55	12.11	6.23

Table C.6.: Systematic uncertainties for normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{d\Delta R_{b_1 b_2}}\right)$ in events with ≥ 3 b-jets.

p_{T,b_1b_2} [GeV]	(0 - 65)	(65 - 100)	(100 - 141)	(140 - 1000)
Measured Value [GeV^{-1}]	0.0043	0.0069	0.0056	0.0003
Powheg+Pythia8 [GeV^{-1}]	0.0040	0.0067	0.0058	0.0003
Uncertainty	%	%	%	%
Data statistics	4.75	5.53	5.70	4.61
MC stats.	6.26	5.21	4.95	8.46
c-jet shape variation	-1.20	-0.31	-0.90	2.38
JES	2.21	1.47	1.91	1.51
JER	0.63	0.78	0.81	0.77
b-tagging calibration	4.99	1.84	2.13	2.38
lepton (reco+ID+iso.)	1.58	0.54	0.64	0.53
lepton trigger	0.88	0.28	0.35	0.29
JVT	0.44	0.12	0.16	0.17
Pileup reweighting	0.75	0.81	0.64	1.03
non-$t\bar{t}$ bkg	1.43	0.80	0.61	2.17
Tot. detector syst.	4.00	4.80	4.90	6.10
shower	1.68	-7.92	1.30	4.74
generator	3.28	-3.06	-0.98	0.27
$t\bar{t}\,V$	0.49	-0.20	-0.14	-0.17
$t\bar{t}\,H$	0.59	-0.07	-0.20	-0.32
ISR/FSR	1.43	4.08	0.80	-5.17
PDF	3.85	1.59	2.18	0.45
Tot. modeling syst.	3.85	9.56	1.96	6.95
Total syst.	5.85	10.57	4.91	9.38
Total	7.52	11.93	7.52	10.45

Table C.7.: Systematic uncertainties on normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dp_{T,b_1b_2}}\right)$ in events with ≥ 3 b-jets.

$m_{b_1 b_2}$ [GeV]	(0 - 50)	(50 - 100)	(100 - 200)	(200 - 400)	(400 - 2000)
Measured Value [GeV^{-1}]	0.0005	0.0027	0.0035	0.0018	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0004	0.0024	0.0036	0.0018	0.0001
Uncertainty	%	%	%	%	%
Data statistics	20.32	7.35	4.29	3.98	6.78
MC stats.	8.96	5.41	5.63	5.54	10.20
c-jet shape variation	-6.07	-1.67	-1.26	1.66	1.56
JES	6.89	2.24	1.83	1.50	3.02
JER	1.57	0.78	0.88	0.80	1.16
b-tagging calibration	6.46	2.73	1.10	1.00	3.21
lepton (reco+ID+iso.)	0.16	0.10	0.09	0.09	0.08
lepton trigger	0.08	0.00	0.00	0.00	0.00
JVT	0.09	0.02	0.01	0.02	0.02
Pileup reweighting	3.30	2.83	1.29	1.43	1.46
non-$t\bar{t}$ bkg	1.46	2.24	0.89	0.14	1.27
Tot. detector syst.	19.80	8.10	4.40	4.40	7.50
shower	-19.31	7.49	-9.19	4.08	11.91
generator	-2.22	5.83	-10.36	7.45	4.08
$t\bar{t}\,V$	-0.14	0.03	-0.02	0.01	-0.00
$t\bar{t}\,H$	0.17	0.28	-0.11	-0.02	0.07
ISR/FSR	13.34	-0.33	1.17	-0.13	-4.28
PDF	0.32	0.75	0.46	0.07	1.83
Tot. modeling syst.	22.66	9.51	13.92	8.48	13.40
Total syst.	30.09	12.49	14.60	9.56	15.36
Total	36.31	14.49	15.22	10.36	16.79

Table C.8.: Systematic uncertainties on normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dm_{b_1 b_2}}\right)$ in events with ≥ 3 b-jets.

$\Delta R_{bb}^{\Delta\text{min}}$	0.4 - 0.7	0.7 - 1.0	1.0 - 1.3	$\geq$ 1.3 - 3.0
Measured Value	0.4404	0.5286	0.5924	0.3127
Powheg+Pythia8	0.4368	0.5241	0.5240	0.3262
Systemtics	%	%	%	%
Data statistics	6.94	6.34	5.89	2.92
MC stats.	5.12	5.74	5.76	6.94
c-jet shape variation	0.60	0.35	0.93	-0.56
JES	3.45	0.86	1.34	1.02
JER	0.11	0.12	0.79	0.23
b-tagging calibration	1.11	0.64	0.85	0.38
lepton (reco+ID+iso.)	0.19	0.08	0.16	0.04
lepton trigger	0.08	0.02	0.02	0.01
JVT	0.01	0.01	0.03	0.01
Pileup reweighting	2.65	0.42	1.29	0.92
non-$t\bar{t}$ bkg	1.10	1.79	1.00	0.65
Tot. detector syst.	7.70	5.10	5.50	2.40
shower	7.71	-12.77	9.60	-0.85
generator	0.29	0.13	10.26	-3.11
$t\bar{t}$ V	-0.05	-0.04	0.00	0.02
$t\bar{t}$ H	-0.22	-0.40	-0.16	0.22
ISR/FSR	2.96	-5.39	8.15	-1.54
PDF	0.52	0.24	0.52	0.22
Tot. modeling syst.	8.28	13.87	16.25	3.59
Total syst.	11.31	14.78	17.16	4.31
Total	13.27	16.08	18.14	5.21

Table C.9.: Systematic uncertainties for normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{d\Delta R_{bb}^{\Delta\text{min}}}\right)$ in events with ≥ 3 b-jets.

$m_{bb}^{\Delta\text{min}}$ [GeV]	(0 - 55)	(55 - 100)	(100 - 145)	(145 - 250)	(250 - 1000)
Measured Value [GeV^{-1}]	0.0042	0.0079	0.0046	0.0015	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0040	0.0076	0.0047	0.0017	0.0001
Uncertainty	%	%	%	%	%
Data statistics	5.74	4.32	5.65	5.94	10.77
MC stats.	5.35	5.64	5.38	10.50	5.59
c-jet shape variation	-0.74	-0.59	0.40	1.92	-0.19
JES	3.46	1.11	2.30	1.58	3.04
JER	1.49	1.16	0.47	0.75	3.24
b-tagging calibration	2.01	0.58	1.00	1.77	3.58
lepton (reco+ID+iso.)	0.08	0.06	0.08	0.16	0.61
lepton trigger	0.02	0.01	0.01	0.00	0.01
JVT	0.02	0.01	0.00	0.01	0.01
Pileup reweighting	2.81	0.83	1.73	2.39	2.20
non-$t\bar{t}$ bkg	0.47	0.63	0.69	1.18	1.32
Tot. detector syst.	7.50	3.60	5.70	6.80	11.10
shower	-5.34	5.29	-5.64	2.63	1.47
generator	-4.98	6.04	-2.29	-7.57	16.48
$t\bar{t}\,V$	-0.08	-0.01	0.06	0.01	0.08
$t\bar{t}\,H$	-0.01	-0.02	-0.25	0.26	0.33
ISR/FSR	2.02	1.89	-0.64	-3.36	-6.50
PDF	1.24	0.32	0.29	1.12	0.47
Tot. modeling syst.	7.67	8.26	6.13	8.76	17.78
Total syst.	10.73	9.01	8.37	11.09	20.96
Total	12.17	9.99	10.10	12.58	23.57

Table C.10.: Systematic uncertainties on normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dm_{bb}^{\Delta\text{min}}}\right)$ in events with ≥ 3 b-jets.

$p_{T,bb}^{\Delta\text{min}}$ [GeV]	(0 - 65)	(65 - 100)	(100 - 150)	(150 - 250)	(250 - 1000)
Measured Value [GeV^{-1}]	0.0015	0.0073	0.0068	0.0024	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0015	0.0072	0.0067	0.0025	0.0001
Uncertainty	%	%	%	%	%
Data statistics	9.79	5.41	4.39	5.01	9.43
MC stats.	4.80	5.02	7.39	5.34	4.37
fit (stat.)	-0.31	-0.23	-0.12	0.46	0.27
c-jet shape variation	-3.81	-1.54	-0.85	3.08	4.54
JES	3.41	1.61	1.03	1.92	2.51
JER	3.56	0.73	1.17	1.08	3.88
b-tagging calibration	7.76	1.52	1.21	1.91	3.52
lepton (reco+ID+iso.)	2.12	0.20	0.31	0.19	0.23
lepton trigger	1.09	0.09	0.14	0.11	0.11
JVT	0.63	0.02	0.08	0.07	0.11
Pileup reweighting	1.69	1.17	0.51	1.22	1.77
non-$t\bar{t}$ bkg	1.30	1.05	0.66	0.88	5.02
Tot. detector syst.	11.50	6.00	4.30	6.50	12.60
shower	-13.65	4.61	-3.87	2.32	14.13
generator	-14.87	6.55	-0.07	-2.53	5.30
$t\bar{t}\,V$	0.54	-0.05	-0.05	-0.09	0.03
$t\bar{t}\,H$	0.86	0.15	-0.00	-0.32	-0.48
ISR/FSR	-24.49	1.95	7.95	-3.06	-0.35
PDF	3.04	0.58	1.06	0.36	1.82
Tot. modeling syst.	32.31	8.28	8.90	4.67	15.32
Total syst.	34.26	9.94	9.68	7.92	20.62
Total	35.61	11.32	10.63	9.38	22.68

Table C.11.: Systematic uncertainties on normalised $\left(\frac{1}{\sigma_{t\bar{t}b}}\frac{d\sigma_{t\bar{t}b}}{dp_{T,bb}^{\Delta\text{min}}}\right)$ in events with ≥ 3 b-jets.

Appendix D.

Normalised Differential Cross-Section for $\geq 4\,b$-jets

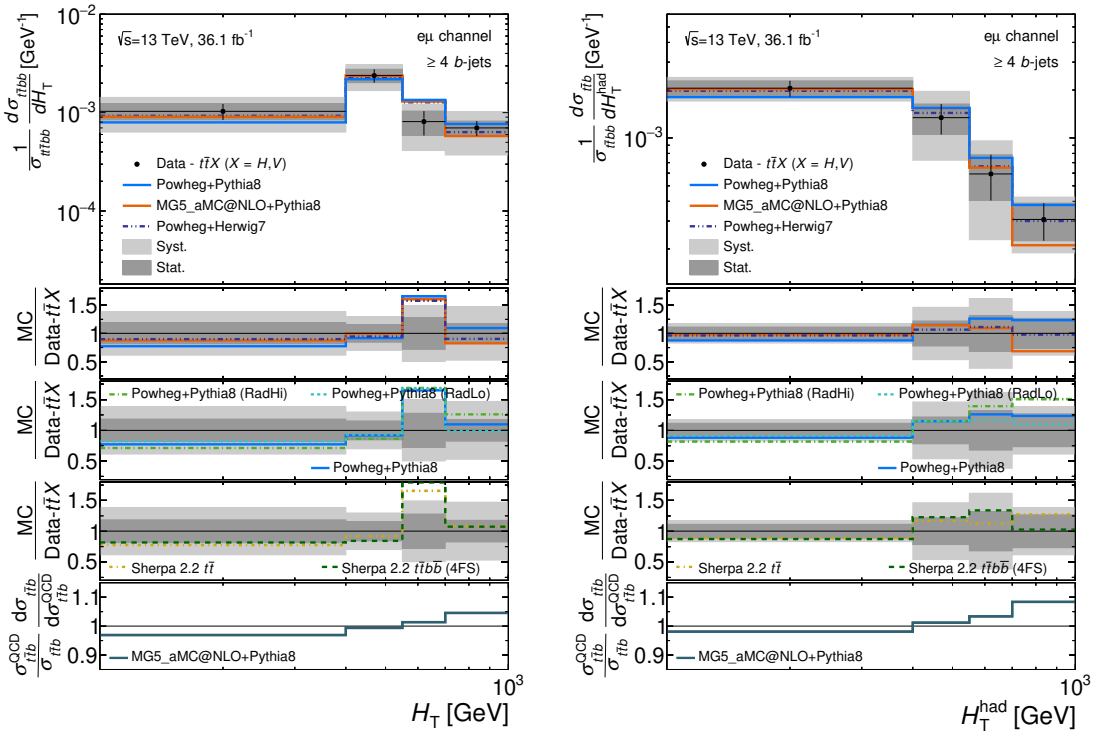


Figure D.1.: Normalised fiducial cross sections as a function of H_{T}, $H_{\mathrm{T}}^{\mathrm{had}}$ in events with at least 4 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}V$ and $t\bar{t}H$ productions.

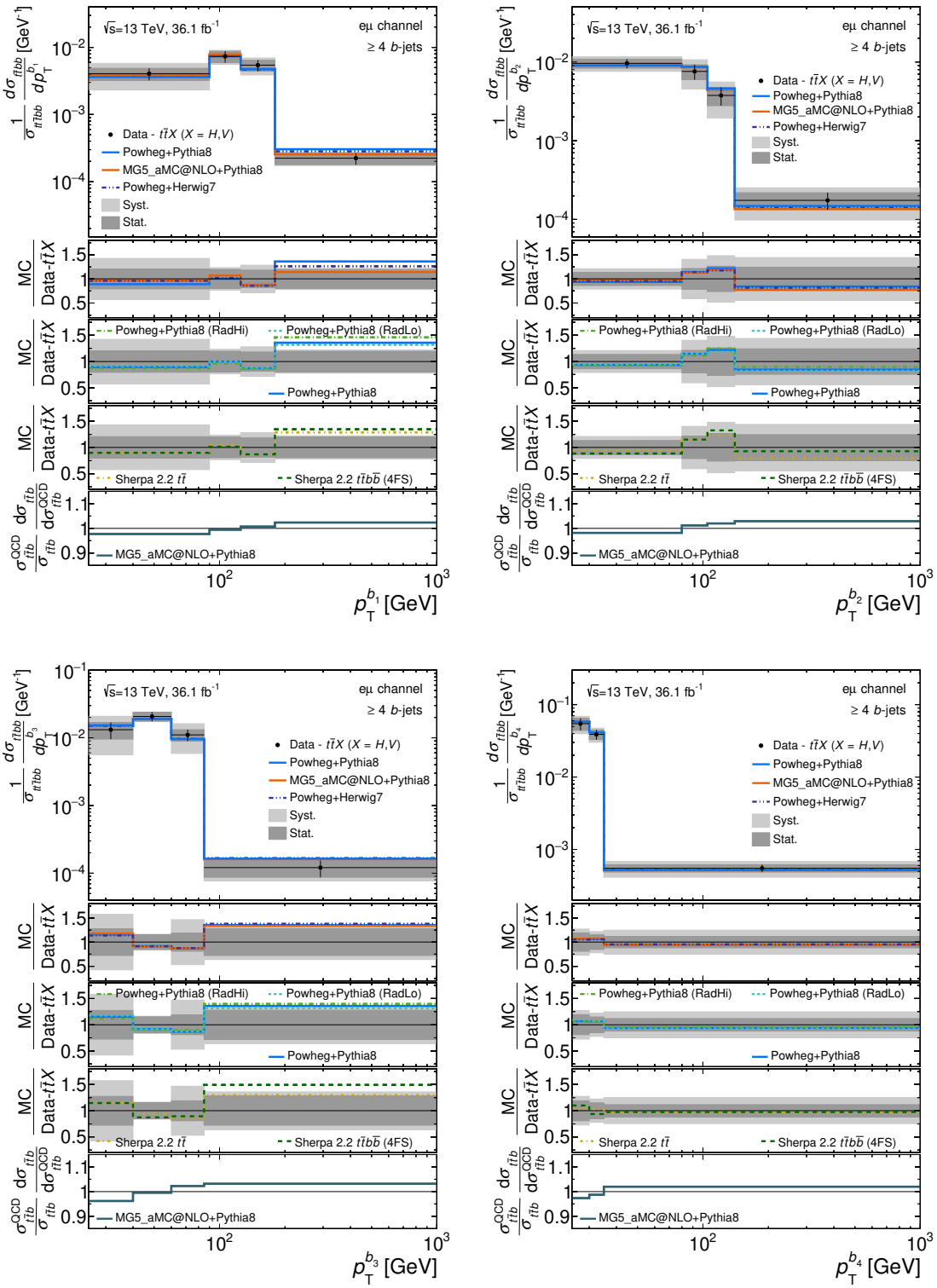


Figure D.2.: Normalised fiducial cross sections as a function of p_{T} for p_{T} ordered jets in events with at least 4 b-jets compared to various MC generators. The third panel shows the ratio of predictions of normalised differential cross-sections from MG5_aMC@NLO+PYTHIA8 including (numerator) and not including (denominator) the contributions of $t\bar{t}V$ and $t\bar{t}H$ productions.

Tables D.1- D.3 show the breakdown of systematic uncertainties of normalised cross section measurements in ≥ 4 b-jets fiducial phase space. Only a few example variables are shown.

$p_{\mathrm{T}}^{b_4}$ [GeV]	(25 - 30)	(30 - 35)	≥ 35
Measured Value [GeV^{-1}]	0.0546	0.0389	0.0006
Powheg+Pythia8	0.0584	0.0415	0.0005
Uncertainty	%	%	%
Data statistics	19.01	15.14	11.81
c-jet shape variation	5.10	2.95	-3.69
JES	3.14	2.43	2.55
JER	0.23	2.12	0.74
b-tagging calibration	5.52	2.88	4.17
lepton (reco+ID+iso.)	0.11	0.06	0.09
lepton trigger	0.01	0.01	0.01
JVT	0.10	0.06	0.08
Pileup reweighting	2.57	2.91	2.65
non-$t\bar{t}$ bkg	8.77	5.16	6.35
Tot. detector syst.	6.20	15.40	18.10
shower	-19.32	-11.93	15.72
generator	-18.53	-4.78	12.26
ttV	-0.07	-0.01	0.04
ttH	-0.91	-0.78	0.79
ISR/FSR	-9.13	-7.42	8.20
PDF	1.24	0.31	0.60
Tot. modeling syst.	28.33	14.87	21.58
Total syst.	29.00	21.41	28.17
Total	34.67	26.22	30.54

Table D.1.: Systematic uncertainties of $\left(\frac{1}{\sigma_{t\bar{t}bb}}\frac{d\sigma_{t\bar{t}bb}}{dp_{\mathrm{T}}^{b_4}}\right)$ in events with ≥ 4 b-jets.

$p_{T,bb}^{\Delta\text{min}}$ [GeV]	(10 - 100)	(100 - 160)	≥ 160
Measured Value [GeV^{-1}]	0.0042	0.0053	0.0004
Powheg+Pythia8 [GeV^{-1}]	0.0036	0.0059	0.0004
Uncertainty	%	%	%
Data statistics	16.44	17.94	15.17
c-jet shape variation	2.08	-1.60	-0.93
JES	2.59	3.79	5.11
JER	1.39	1.80	3.36
b-tagging calibration	4.17	3.97	3.07
lepton (reco+ID+iso.)	0.07	0.21	0.20
lepton trigger	0.01	0.01	0.00
JVT	0.02	0.04	0.02
Pileup reweighting	2.48	4.89	4.46
non-$t\bar{t}$ bkg	6.59	4.84	7.15
Tot. detector syst.	20.70	24.10	13.80
shower	9.71	-9.11	0.35
generator	-15.44	10.37	3.60
ttV	-0.07	0.19	-0.14
ttH	0.11	-0.04	-0.05
ISR/FSR	-11.11	11.06	-1.10
PDF	0.58	0.58	1.18
Tot. modeling syst.	21.37	17.70	3.96
Total syst.	29.75	29.90	14.36
Total	33.99	34.87	20.88

Table D.2.: Systematic uncertainty of $\left(\frac{1}{\sigma_{t\bar{t}bb}}\frac{d\sigma_{t\bar{t}bb}}{dp_{T,bb}^{\Delta\text{min}}}\right)$ in 4 b phase space.

$m_{b_1 b_2}$ [GeV]	(20 - 120)	(120 - 240)	(240 - 400)	(400 - 2000)
Measured Value [GeV^{-1}]	0.0020	0.0033	0.0015	0.0001
Powheg+Pythia8 [GeV^{-1}]	0.0016	0.0031	0.0018	0.0001
Uncertainty	%	%	%	
Data statistics	20.87	15.70	19.03	20.72
c-jet shape variation	2.98	-2.52	0.51	1.56
JES	4.57	2.43	3.28	3.51
JER	3.67	0.12	1.54	0.59
b-tagging calibration	5.59	3.53	3.62	6.01
lepton (reco+ID+iso.)	0.09	0.19	0.07	0.56
lepton trigger	0.01	0.01	0.01	0.01
JVT	0.00	0.01	0.03	0.04
Pileup reweighting	1.95	1.09	1.42	2.65
non-$t\bar{t}$ bkg	2.88	2.45	2.25	3.79
Tot. detector syst.	24.30	10.20	28.90	24.50
shower	-26.11	-10.01	12.77	29.58
generator	-8.10	-23.88	17.26	38.22
ttV	-0.16	0.16	0.01	-0.21
ttH	0.61	-0.64	0.30	0.37
ISR/FSR	-3.74	-8.13	22.84	-15.61
PDF	1.09	0.34	0.19	0.36
Tot. modeling syst.	27.62	27.15	31.35	50.79
Total syst.	36.79	29.00	42.64	56.39
Total	42.30	32.97	46.69	60.08

Table D.3.: Systematic uncertainty of $\left(\frac{1}{\sigma_{t\bar{t}bb}}\frac{d\sigma_{t\bar{t}bb}}{dm_{b_1 b_2}}\right)$ in $\geq 4\,b$ phase space.

Bibliography

[1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1, arXiv:1207.7214 [hep-ex].

[2] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30, arXiv:1207.7235 [hep-ex].

[3] C. Bini, R. Contino, and N. Vignaroli, *Heavy-light decay topologies as a new strategy to discover a heavy gluon*, JHEP **01** (2012) 157, arXiv:1110.6058 [hep-ph].

[4] D. Barducci, A. Belyaev, S. De Curtis, S. Moretti, and G. M. Pruna, *Exploring Drell-Yan signals from the 4D Composite Higgs Model at the LHC*, JHEP **04** (2013) 152, arXiv:1210.2927 [hep-ph].

[5] A. De Simone, O. Matsedonskyi, R. Rattazzi, and A. Wulzer, *A First Top Partner Hunter's Guide*, JHEP **04** (2013) 004, arXiv:1211.5663 [hep-ph].

[6] D. Barducci and C. Delaunay, *Bounding wide composite vector resonances at the LHC*, JHEP **02** (2016) 055, arXiv:1511.01101 [hep-ph].

[7] H. Davoudiasl, J. L. Hewett, and T. G. Rizzo, *Bulk gauge fields in the Randall-Sundrum model*, Phys. Lett. **B473** (2000) 43–49, arXiv:hep-ph/9911262 [hep-ph].

[8] ATLAS Collaboration, M. Aaboud et al., *Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector*, Phys. Lett. **B784** (2018) 173–191, arXiv:1806.00425 [hep-ex].

[9] CMS Collaboration, A. M. Sirunyan et al., *Observation of $t\bar{t}H$ production*, Phys. Rev. Lett. **120** (2018) 231801, arXiv:1804.02610 [hep-ex].

[10] ATLAS Collaboration, M. Aaboud et al., *Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector*, Phys. Lett. **B786** (2018) 59–86, arXiv:1808.08238 [hep-ex].

[11] CMS Collaboration, A. M. Sirunyan et al., *Observation of Higgs boson decay to bottom quarks*, Phys. Rev. Lett. **121** (2018) 121801, arXiv:1808.08242 [hep-ex].

[12] ATLAS Collaboration, *Study of heavy flavor quarks produced in association with top quark pairs at $\sqrt{s} = 7$ TeV using the ATLAS detector*, Phys. Rev. D **89** (2014) 072012, arXiv:1304.6386 [hep-ex].

[13] ATLAS Collaboration, *Measurements of fiducial cross-sections for $t\bar{t}$ production with one or two additional b-jets in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, Eur. Phys. J. C **76** (2016) 11, arXiv:1508.06868 [hep-ex].

[14] CMS Collaboration, *Measurement of $t\bar{t}$ production with additional jet activity, including b quark jets, in the dilepton decay channel using pp collisions at $\sqrt{s} = 8$ TeV*, Eur. Phys. J. C **76** (2016) 379, arXiv:1510.03072 [hep-ex].

[15] CMS Collaboration, *Measurement of the cross section ratio $\sigma_{t\bar{t}b\bar{b}}/\sigma_{t\bar{t}jj}$ in pp collisions at $\sqrt{s} = 8$ TeV*, Phys. Lett. B **746** (2015) 132, arXiv:1411.5621 [hep-ex].

[16] CMS Collaboration, *Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s} = 13$ TeV*, Phys. Lett. B **776** (2018) 355, arXiv:1705.10141 [hep-ex].

[17] ATLAS Collaboration, *Measurements of inclusive and differential fiducial cross-sections of $t\bar{t}$ production with additional heavy-flavour jets in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, JHEP **04** (2019) 046, arXiv:1811.12113 [hep-ex].

[18] S. Hoeche, F. Krauss, P. Maierhoefer, S. Pozzorini, M. Schonherr, and F. Siegert, *Next-to-leading order QCD predictions for top-quark pair production with up to two jets merged with a parton shower*, Phys. Lett. B **748** (2015) 74–78, arXiv:1402.6293 [hep-ph].

[19] Cascioli, Fabio and Maierhöfer, Philipp and Moretti, Niccolo and Pozzorini, Stefano and Siegert, Frank, *NLO matching for $t\bar{t}b\bar{b}$ production with massive b-quarks*, Phys. Lett. B **734** (2014) 210–214, arXiv:1309.5912 [hep-ph].

[20] M. V. Garzelli, A. Kardos, and Z. Trócsányi, *Hadroproduction of $t\bar{t}b\bar{b}$ final states at LHC: predictions at NLO accuracy matched with Parton Shower*, JHEP **03** (2015) 083, arXiv:1408.0266 [hep-ph].

[21] G. Bevilacqua, M. V. Garzelli, and A. Kardos, *$t\bar{t}b\bar{b}$ hadroproduction with massive bottom quarks with PowHel*, arXiv:1709.06915 [hep-ph].

[22] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264–1266.

[23] S. L. Glashow et al., *Weak Interactions with Lepton-Hadron Symmetry*, Phys. Rev. **D2** (1970) 1285–1292.

[24] G. 't Hooft and M. J. G. Veltman, *Regularization and Renormalization of Gauge Fields*, Nucl.

Phys. **B44** (1972) 189–213.

[25] H. D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349, [274(1973)].

[26] D. J. Gross and F. Wilczek, *Asymptotically Free Gauge Theories - I*, Phys. Rev. **D8** (1973) 3633–3652.

[27] E. Noether, *Invariant Variation Problems*, Gott. Nachr. **1918** (1918) 235–257, arXiv:physics/0503066 [physics], [Transp. Theory Statist. Phys.1,186(1971)].

[28] E. Fermi, *On the Quantization of the Monoatomic Ideal Gas*, arXiv e-prints (1999), arXiv:cond-mat/9912229, English Translation by Alberto Zannoni.

[29] P. A. M. Dirac and R. H. Fowler, *On the theory of quantum mechanics*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **112** (1926) 661–677.

[30] W. Pauli, *Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren*, Zeitschrift für Physik **31** (1925) 765–783, https://doi.org/10.1007/BF02980631, English Translation: http://www.fisicafundamental.net/relicario/doc/Pauli_1925.pdf.

[31] Super-Kamiokande Collaboration, Y. Fukuda et al., *Evidence for oscillation of atmospheric neutrinos*, Phys. Rev. Lett. **81** (1998) 1562–1567, arXiv:hep-ex/9807003 [hep-ex].

[32] Particle Data Group Collaboration, M. Tanabashi et al., *Review of Particle Physics*, Phys. Rev. **D98** (2018) 030001.

[33] S. N. Bose, *Planck's Law and Light Quantum Hypothesis*, Zeitschrift für Physik **26** (1924) 178–181, https://doi.org/10.1007/BF01327326, English Version: http://www.entropyofradiation.com/files/Bose/EN_Bose_web.pdf.

[34] A. Einstein, *Quantum Theory of a Monoatomic Ideal Gas*. Sitzungsberichte phys.-math. Kl. Verlag d. Akad. d. Wiss., July, 1924. English Translation by L. Amendola: https://www.thphys.uni-heidelberg.de/~amendola/otherstuff/einstein-paper-v2.pdf.

[35] Eric Drexler, *Elementary particle interactions in the Standard Model*, https://commons.wikimedia.org/wiki/File:Elementary_particle_interactions_in_the_Standard_Model.png.

[36] J. Schwinger, *Quantum Electrodynamics. I. A Covariant Formulation*, Phys. Rev. **74** (1948) 1439–1461, https://link.aps.org/doi/10.1103/PhysRev.74.1439.

[37] R. P. Feynman, *QED: The Strange Theory of Light and Matter*. Prince-ton University Press,

1985. Physics Textbook.

[38] J. G. Korner, A. A. Pivovarov, and K. Schilcher, *On the running electromagnetic coupling constant at M_Z*, Eur. Phys. J. **C9** (1999) 551–556, arXiv:hep-ph/9812465 [hep-ph].

[39] H. Fritzsch, *Quantum Flavordynamics. In: Deeper Pathways in High-Energy Physics*. Springer US, Boston, MA, 1977. Perlmutter A., Scott L.F. (eds).

[40] S. M. Bilenky and J. Hosek, *GLASHOW-WEINBERG-SALAM THEORY OF ELECTROWEAK INTERACTIONS AND THE NEUTRAL CURRENTS*, Phys. Rept. **90** (1982) 73–157.

[41] S. L. Glashow, *Partial Symmetries of Weak Interactions*, Nucl. Phys. **22** (1961) 579–588.

[42] A. Salam, *Renormalizability of gauge theories*, Phys. Rev. **127** (1962) 331–334.

[43] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964) 508–509, [,160(1964)].

[44] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964) 321–323.

[45] F. Mandl and G. Shaw, *Quantum Field Theory*, 2^{nd} *edition*. Wiley , 2010. Physics Textbook.

[46] H. Yukawa, *On the Interaction of Elementary Particles I*, Proc. Phys. Math. Soc. Jap. **17** (1935) 48–57, [Prog. Theor. Phys. Suppl.1,1(1935)].

[47] *Observation of* $\Delta(1234)^{++}$ *production in* e^+e^- *annihilations around 10 GeV*, Physics Letters B **230** (1989) 169 – 174.

[48] R. K. Ellis and W. J. Stirling, *QCD and collider physics*, http://lss.fnal.gov/cgi-bin/find_paper.pl?conf-90-164.

[49] K. G. Wilson, *Confinement of quarks*, Phys. Rev. D **10** (1974) 2445–2459.

[50] M. Thomson, *Modern Particle Physics*. Cambridge University Press, 2013. Physics Textbook.

[51] P. D. Group, *Review of Particle Physics*, Chinese Physics C **40** (2016) 100001.

[52] T. Gleisberg, S. Höche, F. Krauss, M. Schönherr, S. Schumann, et al., *Event generation with SHERPA 1.1*, JHEP **0902** (2009) 007, arXiv:0811.4622 [hep-ph].

[53] D. E. S. John C. Collins and G. Sterman, *Factorization of Hard Processes in QCD*,.

[54] G. Altarelli and G. Parisi, *Asymptotic Freedom in Parton Language*, Nucl. Phys. **B126** (1977) 298–318.

[55] V. N. Gribov and L. N. Lipatov, *Deep inelastic e p scattering in perturbation theory*, Sov. J. Nucl. Phys. **15** (1972) 438–450, [Yad. Fiz.15,781(1972)].

[56] A. D. Martin, *Proton structure, Partons, QCD, DGLAP and beyond*, Acta Phys. Polon. **B39** (2008) 2025–2062, arXiv:0802.0161 [hep-ph].

[57] The ZEUS Collaboration, *Measurement of the neutral current cross section and F_2 structure function for deep inelastic e^+p scattering at HERA*, The European Physical Journal C - Particles and Fields **21** (2001) 443–471.

[58] The H1 Collaboration, *Deep-inelastic inclusive ep scattering at low x and a determination of α_S*, The European Physical Journal C - Particles and Fields **21** (2001) 33–61.

[59] DØ Collaboration, *High-p_T jets in $\overline{p}p$ collisions at $\sqrt{s} = 630$ and 1800 GeV*, Phys. Rev. D **64** (2001) 032003.

[60] CDF Collaboration, *Measurement of the inclusive jet cross section in $p\overline{p}$ collisions at $\sqrt{s} = 1.8$ TeV*, Phys. Rev. D **64** (2001) 032001.

[61] NNPDF Collaboration, R. D. Ball et al., *Parton distributions from high-precision collider data*, Eur. Phys. J. **C77** (2017) 663, arXiv:1706.00428 [hep-ph].

[62] V. V. Sudakov, *Vertex parts at very high-energies in quantum electrodynamics*, Sov. Phys. JETP **3** (1956) 65–71, [Zh. Eksp. Teor. Fiz.30,87(1956)].

[63] F. Hautmann, H. Jung, A. Lelek, V. Radescu, and R. Zlebcik, *Collinear and TMD Quark and Gluon Densities from Parton Branching Solution of QCD Evolution Equations*, JHEP **01** (2018) 070, arXiv:1708.03279 [hep-ph].

[64] LHC Higgs Cross Section Working Group Collaboration, D. de Florian et al., *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*, arXiv:1610.07922.

[65] CDF Collaboration, F. Abe et al., *Observation of top quark production in $\bar{p}p$ collisions*, Phys. Rev. Lett. **74** (1995) 2626–2631, arXiv:hep-ex/9503002 [hep-ex].

[66] D0 Collaboration, S. Abachi et al., *Observation of the top quark*, Phys. Rev. Lett. **74** (1995) 2632–2637, arXiv:hep-ex/9503003 [hep-ex].

[67] M. Cacciari, M. Czakon, M. Mangano, A. Mitov, and P. Nason, *Top-pair production at hadron colliders with next-to-next-to-leading logarithmic soft-gluon resummation*, Phys. Lett. **B710** (2012) 612–622, arXiv:1111.5869 [hep-ph].

[68] M. Beneke, P. Falgari, S. Klein, and C. Schwinn, *Hadronic top-quark pair production with NNLL threshold resummation*, Nucl. Phys. **B855** (2012) 695–741, arXiv:1109.1536.

[69] Bärnreuther, Peter and Czakon, Michal and Mitov, Alexander, *Percent Level Precision Physics at the Tevatron: First Genuine NNLO QCD Corrections to $q\bar{q} \to t\bar{t} + X$*, Phys. Rev. Lett. **109** (2012) 132001, arXiv:1204.5201 [hep-ph].

[70] M. Czakon and A. Mitov, *Top++: A Program for the Calculation of the Top-Pair Cross-Section at Hadron Colliders*, Comput. Phys. Commun. **185** (2014) 2930, arXiv:1112.5675.

[71] LHCTopWG-LHC Top Physics Working Group, 2018.

[72] N. Cabibbo, *Unitary Symmetry and Leptonic Decays*, Phys. Rev. Lett. **10** (1963) 531–533, [,648(1963)].

[73] M. Kobayashi and T. Maskawa, *CP Violation in the Renormalizable Theory of Weak Interaction*, Prog. Theor. Phys. **49** (1973) 652–657.

[74] S. Argyropoulos, J. Katzy, and K. Moenig, *Measurement of the $t\bar{t}b\bar{b}$ production cross-section with 8 TeV ATLAS data*, 2015. https://cds.cern.ch/record/2146901.

[75] F. Maltoni, G. Ridolfi, and M. Ubiali, *b-initiated processes at the LHC: a reappraisal*, JHEP **07** (2012) 022, arXiv:1203.6393 [hep-ph], [Erratum: JHEP04,095(2013)].

[76] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001.

[77] ALICE Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002.

[78] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003.

[79] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004.

[80] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005.

[81] C. Lefèvre, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, https://cds.cern.ch/record/1260465.

[82] ATLAS Collaboration, *ATLAS inner detector: Technical Design Report, 1*, ATLAS-TDR-4 ; CERN-LHCC-97-016, 1997, https://cds.cern.ch/record/331063.

[83] ATLAS Collaboration, *ATLAS Inner Detector Alignment Performance with February 2015 Cosmic Ray Data*, ATL-PHYS-PUB-2015-009, 2015, https://cds.cern.ch/record/2008724.

[84] ATLAS Collaboration, *Study of the material of the ATLAS inner detector for Run 2 of the LHC*, JINST **12** (2017) P12009, arXiv:1707.02826 [hep-ex].

[85] ATLAS Collaboration, *ATLAS Insertable B-Layer Technical Design Report*, 2010,

https://cds.cern.ch/record/1291633.

[86] ATLAS Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*, 1996, https://cds.cern.ch/record/331061.

[87] ATLAS Collaboration, *Readiness of the ATLAS liquid argon calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 723, arXiv:0912.2642 [hep-ex].

[88] ATLAS Collaboration, *Readiness of the ATLAS Tile Calorimeter for LHC collisions*, Eur. Phys. J. C **70** (2010) 1193, arXiv:1007.5423 [hep-ex].

[89] ATLAS Collaboration, *The ATLAS forward calorimeters*, JINST **3** (2008) P02010.

[90] ATLAS Collaboration, *ATLAS muon spectrometer: Technical Design Report*, Geneva, 1997, https://cds.cern.ch/record/331068.

[91] ATLAS Collaboration, *Commissioning of the ATLAS Muon Spectrometer with cosmic rays*, Eur. Phys. J. C **70** (2010) 875, arXiv:1006.4384 [hep-ex].

[92] V. Cindro et al., *The ATLAS beam conditions monitor*, JINST **3** (2008) P02004.

[93] G. Avoni et al., *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*, JINST **13** (2018) P07017.

[94] ATLAS Collaboration, *ATLAS public results of luminosity measurement in Run 2*, https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2.

[95] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector at the LHC*, Eur. Phys. J. C **76** (2016) 653, arXiv:1608.03953 [hep-ex].

[96] N. Bartosik and A. Geiser, *Associated Top-Quark-Pair and b-Jet Production in the Dilepton Channel at $\sqrt{s}$ = 8 TeV as Test of QCD and Background to tt+Higgs Production*, Aug, 2015. https://cds.cern.ch/record/2047049. Presented 03 Jul 2015.

[97] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv:0709.2092 [hep-ph].

[98] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H. S. Shao, T. Stelzer, P. Torrielli, and M. Zaro, *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph].

[99] M. V. Garzelli, A. Kardos, and Z. Trocsanyi, *NLO Event Samples for the LHC*, PoS **EPS-HEP2011** (2011) 282, arXiv:1111.1446 [hep-ph].

[100] G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, A. Kardos, C. G.

Papadopoulos, R. Pittau, and M. Worek, *HELAC-NLO*, Comput. Phys. Commun. **184** (2013) 986–997, arXiv:1110.1499 [hep-ph].

[101] G. Bevilacqua, M. Czakon, M. V. Garzelli, A. van Hameren, Y. Malamos, C. G. Papadopoulos, R. Pittau, and M. Worek, *NLO QCD calculations with HELAC-NLO*, Nucl. Phys. Proc. Suppl. **205-206** (2010) 211–217, arXiv:1007.4918 [hep-ph].

[102] M. Czakon, C. G. Papadopoulos, and M. Worek, *Polarizing the Dipoles*, JHEP **08** (2009) 085, arXiv:0905.0883 [hep-ph].

[103] B. R. Webber, *A QCD Model for Jet Fragmentation Including Soft Gluon Interference*, Nucl. Phys. **B238** (1984) 492–528.

[104] X. Artru and G. Mennessier, *String model and multiproduction*, Nucl. Phys. **B70** (1974) 93–115.

[105] B. Andersson, G. Gustafson, G. Ingelman, and T. Sjostrand, *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31–145.

[106] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *An Introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191** (2015) 159–177, arXiv:1410.3012 [hep-ph].

[107] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, 2014, https://cds.cern.ch/record/1966419.

[108] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82** (2010) 074018, arXiv:1005.3457 [hep-ph].

[109] M. Bahr et al., *Herwig++ Physics and Manual*, Eur. Phys. J. C **58** (2008) 639–707, arXiv:0803.0883 [hep-ph].

[110] B. Johannes et al., *Herwig 7.0/Herwig++ 3.0 release note*, Eur. Phys. J. C **76** (2016) 196, arXiv:1512.01178 [hep-ph].

[111] S. Catani, S. Dittmaier, M. H. Seymour, and Z. Trocsanyi, *The Dipole formalism for next-to-leading order QCD calculations with massive partons*, Nucl. Phys. **B627** (2002) 189–265, arXiv:hep-ph/0201036 [hep-ph].

[112] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **0803** (2008) 038, arXiv:0709.1027 [hep-ph].

[113] *LEP design report*. CERN, Geneva, 1984. https://cds.cern.ch/record/102083. Copies shelved as reports in LEP, PS and SPS libraries.

[114] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*,

JHEP **11** (2004) 040, arXiv:hep-ph/0409146 [hep-ph].

[115] S. Catani, F. Krauss, R. Kuhn, and B. R. Webber, *QCD matrix elements + parton showers*, JHEP **11** (2001) 063, arXiv:hep-ph/0109231 [hep-ph].

[116] L. Lonnblad, *Correcting the color dipole cascade model with fixed order matrix elements*, JHEP **05** (2002) 046, arXiv:hep-ph/0112284 [hep-ph].

[117] M. L. Mangano, M. Moretti, F. Piccinini, and M. Treccani, *Matching matrix elements and shower evolution for top-quark production in hadronic collisions*, JHEP **01** (2007) 013, arXiv:hep-ph/0611129 [hep-ph].

[118] ATLAS Collaboration, *The ATLAS Simulation Infrastructure*, Eur. Phys. J. **C70** (2010) 823–874, arXiv:1005.4568 [physics.ins-det].

[119] Geant4 Collaboration, *GEANT4: A Simulation toolkit*, Nucl. Instrum. Meth. **A506** (2003) 250–303.

[120] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv:1002.2581 [hep-ph].

[121] S. Frixione, P. Nason, and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, JHEP **09** (2007) 126, arXiv:0707.3088 [hep-ph].

[122] D. B. Richard, B. Valerio, C. Stefano, S. D. Christopher, D. D. Luigi, F. Stefano, G. Alberto, P. H. Nathan, I. L. Jose, R. Juan, et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244–289, arXiv:1207.1303 [hep-ph].

[123] NNPDF Collaboration, *Parton distributions for the LHC Run II*, JHEP **04** (2015) 40, arXiv:1410.8849 [hep-ph].

[124] ATLAS Collaboration, *Modelling of the $t\bar{t}H$ and $t\bar{t}V(V = W, Z)$ processes for $\sqrt{s} = 13$ TeV ATLAS analyses*, ATL-PHYS-PUB-2016-005, 2016, https://cds.cern.ch/record/2120826.

[125] ATLAS Collaboration, *Studies on top-quark Monte Carlo modelling for Top2016*, ATL-PHYS-PUB-2016-020, 2016, https://cds.cern.ch/record/2216168.

[126] S. Höche, F. Krauss, M. Schönherr, and F. Siegert, *QCD matrix elements + parton showers: The NLO case*, JHEP **04** (2013) 027, arXiv:1207.5030 [hep-ph].

[127] T. Gleisberg and S. Höche, *Comix, a new matrix element generator*, JHEP **0812** (2008) 039, arXiv:0808.3674 [hep-ph].

[128] F. Cascioli, P. Maierhofer, and S. Pozzorini, *Scattering Amplitudes with Open Loops*, Phys. Rev. Lett. **108** (2012) 111601, arXiv:1111.5206 [hep-ph].

[129] T. Ježo, J. M. Lindert, N. Moretti, and S. Pozzorini, *New NLOPS predictions for $t\bar{t}$ + b-jet production at the LHC*, Eur. Phys. J. C **78** (2018) 502, arXiv:1802.00426 [hep-ph].

[130] N.Kidonakis, *Next-to-next-to-leading-order collinear and soft gluon corrections for t-channel single top quark production*, Phys. Rev. **D 83** (2011), arXiv:1103.2792 [hep-ph].

[131] C. Anastasiou, L. J. Dixon, K. Melnikov, and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at NNLO*, Phys. Rev. **D69** (2004) 094008, arXiv:hep-ph/0312266 [hep-ph].

[132] J. Butterworth et al., *PDF4LHC recommendations for LHC Run II*, J. Phys. **G43** (2016) 023001, arXiv:1510.03865 [hep-ph].

[133] B. Dutta, K. Moenig, T. Lohse, and M. Mozer, *Measurement of W^+W^- production cross section in proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Aug, 2018. https://cds.cern.ch/record/2635975. Presented 19 Jun 2018.

[134] N. A. Asbah and J. Katzy, *Search for the Production of a Standard Model Higgs Boson in Association with Top-Quarks and Decaying into a Pair of Bottom-Quarks with 13 TeV ATLAS Data*, 2018. https://cds.cern.ch/record/2320703. Presented 23 May 2018.

[135] T. Cornelissen, M. Elsing, S. Fleischmann, W. Liebig, E. Moyse, and A. Salzburger, *Concepts, Design and Implementation of the ATLAS New Tracking (NEWT)*, Tech. Rep. ATL-SOFT-PUB-2007-007. ATL-COM-SOFT-2007-002, CERN, Geneva, Mar, 2007. https://cds.cern.ch/record/1020106.

[136] ATLAS Collaboration, *Performance of the ATLAS Track Reconstruction Algorithms in Dense Environments in LHC Run 2*, Eur. Phys. J. **C77** (2017) 673, arXiv:1704.07983 [hep-ex].

[137] ATLAS Collaboration, *Performance of primary vertex reconstruction in proton–proton collisions at $\sqrt{s} = 7$ TeV in the ATLAS experiment*, ATLAS-CONF-2010-069, 2010, https://cds.cern.ch/record/1281344.

[138] ATLAS Collaboration, *Reconstruction of primary vertices at the ATLAS experiment in Run 1 proton-proton collisions at the LHC*, Eur. Phys. J. **C77** (2017) 332, arXiv:1611.10235 [physics.ins-det].

[139] ATLAS collaboration, *Electron reconstruction and identification efficiency measurements with the ATLAS detector using the 2011 LHC proton-proton collision data*, Eur. Phys. J. **C74** (2014) 2941, arXiv:1404.2240 [hep-ex].

[140] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016, https://cds.cern.ch/record/2157687.

[141] W. Lampl et al., *Calorimeter Clustering Algorithms: Description and Performance*, ATL-LARG-PUB-2008-002, 2008, https://cds.cern.ch/record/1099735.

[142] ATLAS Collaboration, *Improved electron reconstruction in ATLAS using the Gaussian Sum Filter-based model for bremsstrahlung*, ATLAS-CONF-2012-047, 2012, https://cds.cern.ch/record/1449796.

[143] ATLAS Collaboration, *Electron reconstruction and identification efficiency measurements in 2016 Data* , https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/EGAM-2017-003/index.html.

[144] ATLAS Collaboration, *Electron and photon energy calibration with the ATLAS detector using data collected in 2015 at* $\sqrt{s} = 13$ *TeV*, ATL-PHYS-PUB-2016-015, 2016, https://cds.cern.ch/record/2203514.

[145] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at* $\sqrt{s}$ *=13 TeV*, Eur. Phys. J. **C76** (2016) 292, arXiv:1603.05598 [hep-ex].

[146] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton–proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76** (2016) 292, arXiv:1603.05598 [hep-ex].

[147] ATLAS Collaboration, *Topological cell clustering in the ATLAS calorimeters and its performance in LHC Run 1*, arXiv:1603.02934 [hep-ex].

[148] M. Cacciari, G. P. Salam, and G. Soyez, *The anti-*k_t *jet clustering algorithm*, JHEP **04** (2008) 063, arXiv:0802.1189 [hep-ph].

[149] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. **D96** (2017) 072002, arXiv:1703.09665 [hep-ex].

[150] ATLAS Collaboration, *JES Public Plots for Moriond 2017* , https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/JETM-2017-003/.

[151] ATLAS Collaboration, *Tagging and suppression of pileup jets with the ATLAS detector*, ATLAS-CONF-2014-018, 2014, https://cds.cern.ch/record/1700870.

[152] ATLAS Collaboration, *Jet energy resolution in proton–proton collisions at* $\sqrt{s} = 7$ *TeV*

recorded in 2010 with the ATLAS detector, Eur. Phys. J. C **73** (2013) 2306, arXiv:1210.6210 [hep-ex].

[153] ATLAS Collaboration, *Data-driven determination of the energy scale and resolution of jets reconstructed in the ATLAS calorimeters using dijet and multijet events at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2015-017, 2015, https://cds.cern.ch/record/2008678.

[154] ATLAS Collaboration, *Determination of the jet energy scale and resolution at ATLAS using* Z/γ*-jet events in data at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2015-057, 2015, https://cds.cern.ch/record/2059846.

[155] ATLAS Collaboration, *Optimisation of the ATLAS b-tagging performance for the 2016 LHC Run*, ATL-PHYS-PUB-2016-012, 2016, https://cds.cern.ch/record/2160731.

[156] ATLAS Collaboration, *Commissioning of the ATLAS high performance b-tagging algorithms in the 7 TeV collision data*, ATLAS-CONF-2011-102, 2011, https://cds.cern.ch/record/1369219.

[157] G. Piacquadio and C. Weiser, *A new inclusive secondary vertex algorithm for b-jet tagging in ATLAS*, J. Phys. Conf. Ser. **119** (2008) 032032.

[158] R. Fruhwirth, *Application of Kalman filtering to track and vertex fitting*, Nucl. Instrum. Meth. **A262** (1987) 444–450.

[159] ATLAS Collaboration, *Measurements of b-jet tagging efficiency with the ATLAS detector using* $t\bar{t}$ *events at* $\sqrt{s} = 13$ *TeV*, JHEP **08** (2018) 089, arXiv:1805.01845 [hep-ex].

[160] ATLAS Collaboration, *Calibration of b-tagging using dileptonic top pair events in a combinatorial likelihood approach with the ATLAS experiment*, ATLAS-CONF-2014-004, 2014, https://cds.cern.ch/record/1664335.

[161] ATLAS Collaboration, *b-tagging calibration plots using dileptonic* $t\bar{t}$ *events produced in* pp *collisions at* $\sqrt{s} = 13$ *TeV and a combinatorial likelihood approach*, http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PLOTS/FTAG-2016-003/.

[162] ATLAS Collaboration, *Performance of b-Jet Identification in the ATLAS Experiment*, JINST **11** (2016) P04008, arXiv:1512.01094 [hep-ex].

[163] ATLAS Collaboration, *Calibration of light-flavour jet b-tagging rates on ATLAS proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, http://cds.cern.ch/record/2314418.

[164] M. Cacciari, G. P. Salam, and G. Soyez, *The Catchment Area of Jets*, JHEP **04** (2008) 005, arXiv:0802.1188 [hep-ph].

[165] ATLAS Collaboration, *Measurement of the* $t\bar{t}$ *production cross-section using* $e\mu$ *events with*

b-tagged jets in pp collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS detector, Eur. Phys. J. C **74** (2014) 3109, arXiv:1406.5375 [hep-ex].

[166] S. Frixione, E. Laenen, P. Motylinski, B. Webber, and C. D. White, *Single-top hadroproduction in association with a W boson*, **07** (2008) 029, arXiv:0805.3067 [hep-ph].

[167] N. Kidonakis, *Two-loop soft anomalous dimensions for single top quark associated production with a W or H*, Phys. Rev. D **82** (2010) 054018, arXiv:1005.4451 [hep-ph].

[168] ATLAS Collaboration, *Measurement of the $t\bar{t}$ production cross-section using $e\mu$ events with b-tagged jets in pp collisions at $\sqrt{s}$=13 TeV with the ATLAS detector*, Phys. Lett. **B761** (2016) 136, arXiv:1606.02699 [hep-ex].

[169] ATLAS Collaboration, M. Aaboud et al., *Measurement of jet activity produced in top-quark events with an electron, a muon and two b-tagged jets in the final state in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, Eur. Phys. J. **C77** (2017) 220, arXiv:1610.09978 [hep-ex].

[170] B. Wynne, *Imagiro: an implementation of Bayesian iterative unfolding for high energy physics*, arXiv:1203.4981 [physics.data-an].

[171] G. D'Agostini, *A Multidimensional unfolding method based on Bayes' theorem*, Nucl. Instrum. Meth. A **362** (1995) 487–498.

[172] T. Adye, *Unfolding algorithms and tests using RooUnfold*, arXiv:1105.1160 [physics.data-an].

[173] *Measurements of $t\bar{t}$ cross sections in association with b jets and inclusive jets and their ratio using dilepton final states in pp collisions at $\sqrt{s}$=13TeV*, Physics Letters B **776** (2018) 355 – 378, http://www.sciencedirect.com/science/article/pii/S0370269317309358.

List of figures

List of tables